UNIVERSIDAD AUTÓNOMA DE MADRID

Facultad de Ciencias

Departamento de Biología

Characterization of a partner switching system regulating c-di-GMP levels in *Sinorhizobium meliloti*. Implication in the synthesis of a novel exopolysaccharide.

Irene Baena Ropero

Madrid, 2015

Memoria para optar al Título de Doctor, realizada en el Departamento de Biología de la Facultad de Ciencias de la Universidad Autónoma de Madrid, y que lleva por título:

Characterization of a partner switching system regulating c-di-GMP levels in *Sinorhizobium meliloti*. Implication in the synthesis of a novel exopolysaccharide.

Presentada por Irene Baena Ropero, Licenciada en Biología.

Dirigida por:

Ildefonso Bonilla Mangas
Catedrático de Universidad
Departamento de Biología

Fco. Javier Lloret Romero
Profesor Contratado Doctor
Departamento de Biología

Madrid, 2015

Ildefonso Bonilla Mangas, Catedrático de Universidad, y Fco. Javier Lloret Romero, Profesor Contratado Doctor, ambos adscritos al Departamento de Biología de la Facultad de Ciencias de la UAM, certifican:

que la memoria para optar al Título de Doctor presentada por la Licenciada en Biología Irene Baena Ropero ha sido realizada bajo su dirección.

Fdo. Ildefonso Bonilla Mangas **Fdo. Fco. Javier Lloret Romero**

AGRADECIMIENTOS

Ya cuando se vislumbraba, muy muy a lo lejos, el final de esta tesis, empecé a pensar en todos los "y si…" y los puntos de inflexión que han ido orientando el camino en el que ahora me encuentro. Unas veces el azar, y otras la buena voluntad de unas cuantas personas, han sido fundamentales para avanzar. El comienzo pudo ser cuando, gracias a María Jesús, me animé a pedir una beca para el Máster que iba a empezar ese curso. Necesitaba encontrar un tutor en tiempo récord. Después de dos puertas cerradas, llegué de rebote al despacho de Rafa, que a su vez me presentó a Javi. Resultó que los dos éramos unos novatos, pero alguna vez tiene que ser la primera, ¿no? Al final, con mucho tiempo y esfuerzo, lo hemos conseguido Javi. Gracias por ser tan cercano, por enseñarme, guiarme, y por tener siempre una sonrisa en la cara. En ésta última recta final han sido muy importantes los puntos de vista externos para mejorar la comprensión y la estructura del texto. Gracias a María Sánchez y a María Jesús por vuestras correcciones y sugerencias. A Miguel, por ayudarme con los análisis y solucionarme todos los problemas informáticos, yo que tú montaba un negocio en paralelo, te tenemos explotado. A Cris, por ayudarme con los últimos experimentos de proteínas, dejarme los anticuerpos, y enseñarme los protocolos.

Como suele ser ley de vida, los becarios veteranos y postdoc acaban enseñando muchas, muuuchas cosas a los nuevos. Ahí estaban Juan, María Reguera, Ángel, Diego, Sandra, Belén, Laura, Ana, Emma, Fran, Samu, Virginia, Carol, Cristina, David, Ramsy…Gracias especialmente a Ana, Emma y Juan por acogerme desde el principio, por vuestra paciencia y generosidad, y por ayudarme con todas las dudas que me surgían. También a Fran, que cogió el relevo más adelante y al que también di bastante la lata. A Ángel, por su sentido del humor y las tardes sabineras que hacían todo más llevadero.

Conseguir una beca no resultó fácil, y en ese periodo hubo uno de esos puntos de inflexión importantes. Gracias a Alfonso, que a pesar de no tenerle tan informado como hubiera debido, siempre me ayudó con los temas burocráticos, con sus consejos, y además me dio la oportunidad de continuar un año más, lo justo para poder pedir por tercera y última vez la beca que al final conseguí. Así que probablemente no estaría aquí de no ser por él. Muchas, muchas gracias por confiar en mí, y darme esa oportunidad.

Después de los resultados "negativos" del Máster, llegaron los resultados raros, feos, de esos que hacen que te des cabezazos con la pared mientras piensas "por qué me metí en

esto". Y entonces apareció el "y si…" de Javi. ¿Y si mutamos el gen de al lado? Y así, apareció el extremo del hilo de la madeja. Más adelante yo le diría "y si vamos a este congreso que tiene buena pinta? Y allí coincidimos con María José Soto, que tuvo el gran gesto (muchas gracias) de animarnos a hablar con Juan Sanjúan, porque le parecía que estábamos haciendo cosas parecidas. Otro punto de inflexión. ¿Y si no hubiéramos coincidido en el congreso? ¿Y si no hubiera sido Juan, sino un investigador desconocido, sin escrúpulos, y de las antípodas? De manera que me fui para Granada. Y allí la gente me trató de maravilla. Gracias a Lorena y Nieves (¡ánimo con esas tesis!), a David (qué increíble tener un técnico), a Lydia, a Paola. Muy especialmente a Dani, por su infinita paciencia, su sentido del humor, y por todas las ideas que tanto él como Juan Sanjuán han aportado a este proyecto.

A la comisión de Fisiología Vegetal, muchas gracias por acogerme, enseñarme, y ayudarme. Sois un cuerpazo docente, no dejo de asombrarme de vuestra capacidad para gestionar tantas cosas a la vez, y aun así tener tiempo para celebrar cualquier buena noticia. Y es que a lo largo de estos años, no sólo he aprendido a investigar, sino también a enseñar. Son varios los estudiantes que han pasado por el laboratorio: Carlos, Silvia, Elena, María, Irene, Kevin. Gracias a todos por los datos que habéis aportado y por vuestra paciencia con esta novata. No sé si os habré ayudado lo suficiente, pero yo desde luego he aprendido mucho. Ahora que lo veo desde "el otro lado", me doy cuenta de lo difícil que es enseñar, y admiro a aquellos que realmente lo consiguen.

A mis compañeros predoctorales o pretésicos: Andrea, Candela, Maiki, Keila, Gerardo, con los que he compartido desahogos, prácticas, tareas, organización, y también me he reído mucho (viva la cervecita de los viernes, un bien muy necesario). Mucho ánimo a todos, ¡todo llega! Un agradecimiento muy especial para Isidro, mi compañero de beca, de politiqueo, de burocracia, de lotería, de cervecitas (y probablemente de tartazo). De mayor quiero ser como tú. Eres una persona (y un investigador) brillante. Siempre encuentras hueco para escuchar y ayudar a la gente, eres paciente, y da gusto hablar de cualquier tema contigo. Acuérdate de mí cuando seas jefazo, gracias.

A Claude Bruand, por acogerme durante mi estancia en Toulouse y solucionarme cualquier problema que surgiera. Gracias especialmente a Laurent Sauviac, que estuvo conmigo codo con codo intentando sobreexpresar mis proteínas e incluso construyó tres plásmidos nuevos. No pudo ser, pero de todo se aprende.

A las nuevas incorporaciones y generaciones de becarios: Laura, Jara, Dani, Eva, y los que tuvieron que marchar a buscarse la vida, Pili y Narciso. Sois estupendos, os deseo mucha suerte, aprended, y disfrutad, que al final se pasa volado. Creo que nunca me he reído tanto en las comidas como con vosotros, viva el humor y las conversaciones absurdas.

Gracias a mis padres, que al final son los principales responsables de ser como soy, por la educación que me han dado y los valores que he aprendido de ellos. Siempre me han apoyado para hacer lo que realmente me gustara, aunque con el paso del tiempo mis palabras les fueran sonando cada vez más a un idioma extraterrestre.

No podía olvidarme de la **B**iología, el punto de inflexión más importante de mi vida hasta ahora. De no ser por ella, puede que no me hubiera cruzado contigo, Chus, mi mejor apoyo, mi mejor todo, mi impronta. GRACIAS.

RESUMEN

Sinorhizobium meliloti produce un β-D-glucano de enlaces alternos (1→3) (1→4) (ML β-glucano) en respuesta a altos niveles de diguanilato cíclico (di-GMP-c). Dos proteínas, BgsB y BgsA, son las responsables de la síntesis, siendo BgsA la glucano sintasa que sensa dichos niveles de di-GMP-c. La transcripción de los genes *bgsBA* depende del regulador global ExpR, perteneciente al sistema de quorum sensing Sin/ExpR. Este estudio se centra en la caracterización de un operón que regula la síntesis del ML β-glucano en *S. meliloti*. Dicho operón fue identificado a raíz de un análisis fenotípico de mutantes afectados en el metabolismo del di-GMP-c. De los seis genes que lo constituyen, *SMb20447* codifica para una proteína anotada como diguanilato ciclasa/fosfodiesterasa. Se demostró que esta proteína es activa como diguanilato ciclasa y que induce la síntesis del ML β-glucano.

A diferencia de lo observado con los genes *bgsBA,* la transcripción del operón de estudio no es dependiente de ExpR, formando así un sistema de regulación adicional. Al menos tres de las proteínas codificadas en este operón conforman un sistema de "partner switching" semejante al que existe en *Bacillus subtilis*, el cual se encarga de regular respuestas frente a stress a través de los factores de transcripción σ^B y σ^F. Mediante un abordaje genético, hemos determinado que las proteínas SMb20450 y SMb20451, que presentan dominios característicos de fosfatasas y serin-quinasas respectivamente, modulan la actividad represora de SMb20452, una proteína con dominio STAS presente en factores anti-anti-sigma y en transportadores de sulfato. Este sistema regula la actividad diguanilato ciclasa de SMb20447 a nivel posttranscripcional, a través de una posible interacción con SMb20452. El operón parece estar bien conservado en especies de la familia de las Rizobiáceas que también presentan los genes *bgsBA*, por lo que el sistema de regulación del ML β-glucano en estas bacterias podría ser parecido.

En cuanto a su papel biológico, el ML β-glucano es necesario durante el proceso de adhesión a las raíces de alfalfa. Sin embargo, su producción de forma constitutiva no supone ninguna ventaja en relación a este tipo de adhesión. El ML β-glucano no participa en la interacción simbiótica con la planta, ya que no sustituye las funciones desempeñadas por los exopolisacáridos succinoglicano (EPS I) o galactoglucano (EPS II). Un análisis transcriptómico global mostró que el operón *SMb20447-SMb20452* podría regular otros procesos como la producción de EPS II y sideróforos, adaptación a condiciones microaeróbicas, o transporte de cationes.

SUMMARY

Sinorhizobium meliloti synthesizes a mixed-linked (1→3) (1→4)-β-D-glucan (ML β-glucan) in response to high levels of cyclic diguanylate (c-di-GMP). Two proteins, BgsA and BgsB, are required for the synthesis, being BgsA the glucan synthase sensing c-di-GMP levels. The transcription of the *bgsBA* operon is dependent on the global regulator ExpR, which also forms part of the quorum sensing system Sin/ExpR. This study is focused on the characterization of an operon regulating the synthesis of the ML-β glucan in *S. meliloti*. The operon was identified as a consequence of a screening of mutants affected in c-di-GMP metabolism. Among the six genes that constitute the operon, *SMb20447* codes for a protein annotated as a diguanylate cyclase/phosphodiesterase. We demonstrated this protein is active as a diguanylate cyclase, and triggers the synthesis of the ML β-glucan.

Unlike the *bgsBA* genes, the transcription of the operon is not dependent on ExpR, thus forming and additional regulatory system. At least three of the gene products in the operon seem to form a partner switching system that resembles the one regulating *Bacillus subtilis* general stress response through sigma factors σ^B and σ^F. Using a genetic approach we determined that the proteins SMb20450 and SMb20451, which present putative phosphatase and kinase effector domains respectively, modulate the repressor activity of SMb20452, a STAS (sulphate transporter and anti-sigma antagonist) domain protein. The system regulates the diguanylate cyclase activity of SMb20447 at the posttranscriptional level, probably through a direct interaction with SMb20452. The operon is well conserved in bacteria from the Rhizobiaceae family that present the *bgsBA* genes, indicating a similar role in these putative ML β-glucan producers.

Regarding its biological role, the ML β-glucan proved to be very important for the attachment to alfalfa roots. However, its constitutive expression does not provide any advantage in relation to such attachment. This novel EPS did not present a symbiotic function, since it could not substitute either succynoglican (EPS I) nor galactoglucan (EPS II). Global transcriptomic analysis revealed that the *SMb20447-SMb20452* operon might regulate other processes like galactoglucan (EPS II) and siderophore production, adaptation to microaerobic environments, or cation transport.

INDEX

KEY ABBREVIATIONS

aa = amino acid
Amp = ampicillin
ANOVA = Analysis of variance
bp = base pair
c-di-GMP = bis-(3'-5')-cyclic dimeric guanosine monophosphate
CF = Calcofluor
CG = cyclic β-glucan
CPS = capsular polysaccharide
CR = Congo red
DGC = diguanylate cyclase
EPS = exopolysaccharide
FTIR = Fourier Transform Infrared Spectroscopy
Gn = gentamycin
GPS = gel forming polysaccharide
kb = kilo base
Kn = kanamycin
KPS = k-antigen polysaccharide
MFS = Major Facilitator Superfamily
MM = Minimal Medium
Nm = neomycin
Nod factor = nodulation factor
NP = neutral polysaccharide
OD = Optical Density
ORF = Open Reading Frame
PDE = phosphodiesterase
Str = streptomycin
Tmp = trimethoprim
TY = Tryptone –Yeast
UV = ultraviolet

INTRODUCTION

The Rhizobia group.

Rhizobia are a large group of gram-negative soil bacteria characterized by their ability to stablish symbiotic relationships with legume plants. It comprises five families of α-proteobacteria: Rhizobiaceae (*Rhizobium, Sinorhizobium* or *Ensifer, Allorhizobium, Shinella*), Phyllobacteriaceae (*Mesorhizobium, Aminobacter, Phylobacterium*), Bradyrhizobiaceae (*Bradyrhizobium*), Methylobacteriaceae (*Methylobacterium, Microvirga*), and Hyphomicrobiaceae (*Azorhizobium, Devosia*) There are also several genera from the β-proteobacteria group: *Burkholderia*, *Ralstonia, Cupriavidus, Herbaspirillum* (Vinuesa, 2015). Generally, Rhizobia induce the formation of specific organs in the root or the stem of the plant known as nodules, where bacteria are able to fix atmospheric nitrogen and produce ammonium, which is then assimilated by the legume (Jones *et al.*, 2007, Gibson *et al.*, 2008). This capacity means a great advantage for plants that live in nitrogen-poor soils, and has given rise to a very interesting research field in relation to the application of this type of bacteria as natural fertilizers (Miransari, 2011, Adesemoye *et al.*, 2009, Lindstrom *et al.*, 2010).

The establishment of this symbiosis is based on a very complex signal dialog between rhizobia and the legume plant. This dialog begins when aromatic compounds called flavonoids secreted by the plant roots are recognized by the bacteria, which in response produce the Nodulation factors responsible for the induction of the nodule development in the plant (Gibson *et al.*, 2008). The structure of Nod factors consist of a backbone of N-acetylglucosamine oligomer with β (1→4) linkages, a fatty acid attached to the terminal glucosamine and a different set of substitutions along the mentioned backbone. These substitutions sometimes influence in the specificity of the bacteria-legume interaction, so that for a given Rhizobia species, only a particular group of legumes will be successfully nodulated (Hirsch *et al.*, 2001). Nod factors act as diffusible signals that trigger different processes in the epidermis and the cortex of the plant root. They produce alterations in the cytoskeleton of the cells from the root hairs tips, provoking its bending and enclosing the bacteria as a consequence. Then, the root hair tip wall experiences an invagination, forming a tubule filled with bacteria (the infection thread) that elongates towards the inner plant cortex, where bacteria will be eventually engulfed by the plant cells. Nod factors stimulate the mitotic division of the cortex cells, which will form the so-called "nodule primordium". The bacterial cells are endocytosed individually by the cortex cells, inside unwalled membrane compartments originated from the infection thread. This set of bacterial cell and endocytosis membrane is known as symbiosome. Eventually, rhizobia cells will

differentiate into the nitrogen-fixing bacteroid form (Jones *et al.*, 2007, Gage, 2004). A general view of this symbiotic process is described in Figure 1.

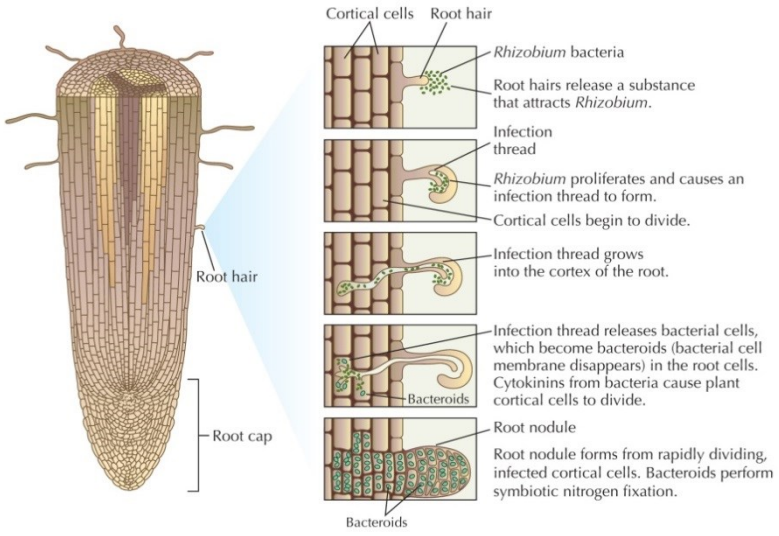


Figure 1. Overview of the symbiotic process between Rhizobia and legume plants. Figure reproduced from *Evolution*© 2007 Cold Spring Harbor Laboratory Press.

Two different types of nodules have been described: indeterminate and determinate. The formers are elongated and present a persistent meristem in the inner root cortex that continuously provides new nodule cells which will be infected by incoming rhizobia from the infection threads or already present in the nodule. The new infected cells will develop more nodule tissue with nitrogen-fixing capacity. Thus, indeterminate nodules display a gradient of developmental stages, from the uninfected meristem cells at the tip of the nodule, to the older senescent tissue located near the root. Legumes like *Medicago sativa* (alfalfa), *Pisum sativum* (pea), *Vicia sp.* (vetches) or *Trifolium sp.*(clovers) have been traditionally used as models for the study of indeterminate nodules development (Gage, 2004, Gibson *et al.*, 2008). In contrast, determinate nodules are round, with a temporal meristem initiated in the outer root cortex. Rhizobia cells mainly expand thanks to the division of already infected cells from the outer cortex, so generally all cells in the nodule present the same developmental stage. Some examples of legumes that forms determinate nodules are *Glycine max* (soybean), *Vicia faba* (bean), and *Lotus japonicus* (Gage, 2004, Gibson *et al.*, 2008).

Bacterial exopolysaccharides and biofilms.

Bacterial exopolysaccharides (EPS) play an important role in the free-living stage of the cells, but also during infective or symbiotic processes. They form part of the complex outer surface of the cell, together with lipopolysaccharides (LPS), capsular (CPS) and K-antigen (KPS) polysaccharides, cyclic β-glucans (CG), gel-forming polysaccharides (GPS), or neutral polysaccharides (NP or glucomannan) (Skorupska *et al.*, 2006, Janczarek, 2011). Generally, EPS are weakly associated to the bacterial surface and are released in large amounts to the environment. They are linear or branched heteropolymers and homopolymers composed of a repeating unit of different common monosaccharides like D-glucose, D-galactose, L-rhamnose, D-glucuronic acid, D-mannose, etc. The repeating unit is often species-specific and can present non-carbohydrate substitutions (acetyl, pyruvyl, succynil groups) that confer the EPS an acidic nature. EPS can also present different polymerization degrees (high or low molecular-weight forms), which will be crucial for some physiological processes. Thus, there are many possible combinations regarding the type of monosaccharides, non-carbohydrate substitutions, glucosidic linkages, degree of polymerization and so on that make the EPS a very diverse group (Janczarek, 2011, Skorupska *et al.*, 2006). It is worthy to distinguish that minor modifications in the structure of this polymers can change drastically their chemical and biological properties (Schmid *et al.*, 2015, Whitfield *et al.*, 2015).

Exopolysaccharides constitute the main structure of mature bacterial biofilms. Biofilms are defined as bacterial communities surrounded by a self-produced polymeric matrix that can reversibly get attached to either inert or biotic surfaces (Costerton *et al.*, 1995). They suppose a major way of growth in nature, likely due to their protective character, and have been well described in a wide range of bacteria, from human and plant pathogens such as species from genera *Salmonella, Yersinia*, *Staphyloccocus, Xhantomonas*, to non-pathogen ones like those from genera *Rhizobium, Vibrio,* or some *Pseudomonads* (Ramirez-Mata *et al.*, 2014).

Water is the main component of biofilms, followed by cells and EPS. Other minor components, but not less important, are lipopolysaccharides (LPS), extracellular DNA (eDNA), some enzymes, and proteins (Sutherland, 2001). In order to assure the survival of the cells, mature biofilms are often crossed by a sophisticated network of channels that facilitates the diffusion of nutrients and other substances (Stanley & Lazazzera, 2004). Two different phases can be differentiated during biofilm formation: an initial, non-specific, and reversible stage mediated by hydrophobic and electrostatic interactions between the free-

living cells and the adjacent surfaces; and an irreversible stage where the adhesion process is completed and bacterial micro colonies are stablished. Several components such as flagella, pili, fimbriae and LPSs are crucial along the initial stage, whereas EPS are the scaffolds of mature biofilms, providing mechanical stability and cohesive interactions with the interface. Under certain conditions, cells from the mature biofilm will eventually detach from the matrix and revert to the free-living and motile form, ready to search for new niches (Bogino *et al.*, 2013, Stanley & Lazazzera, 2004). Thus, EPS together with the rest of the matrix components prevent the diffusion of antibiotics, defense substances, and offer protection against different environmental stresses like desiccation, nutrient starvation, pH changes, ultraviolet light radiation, osmotic changes, protozoan predators, etc. (Flemming & Wingender, 2010).

As outlined before, the chemical structure of the EPS will define its biological properties. Most EPS have a polyanionic nature, for instance, due to the presence of uronic acids and sugars with pyruvate, phosphate or sulfate substitutions. This allows the EPS to stablish unions with divalent cations like Ca^{2+} or Mg^{+2} that improve the stability of the matrix. On the other hand, β (1→3) or β (1→4) linkages confer more rigidity than the α (1→2) or α (1→6) ones (Bogino *et al.*, 2013). This type of properties will determine the interactions of the EPS with its surroundings, as well as the stability of the biofilm matrix. In some cases, the great adaptation skills of bacteria allow them to produce more the one type of EPS depending on the environmental conditions, as it has been demonstrated for *P. aeruginosa* and *Streptococcus termophilus* (Ryder *et al.*, 2007, Vaningelgem *et al.*, 2004). The main carbon source for growth can also modify the composition of the EPS, as it was described for *Bradyrhizobium japonicum* (Karr *et al.*, 2000).

Rhizobial EPS. Biological role and applications.

As outlined before, EPS can be very diverse regarding chemical structure and properties. In relation to Rhizobia, a small sample of this variability is shown in Figure 2. Rhizobial exopolysaccharides have been mainly studied in relation to the symbiotic process. In indeterminate nodule-legumes, acidic EPS deficient strains from *S .meliloti* and *R. leguminosarum* fail to develop complete infection threads (Leigh *et al.*, 1985, Breedveld *et al.*, 1993). As a consequence, bacteria are not able to reach the cortical cells, and the emerging nodules do not fix nitrogen. There are some hypotheses suggesting that EPS, like

succinoglycan from *S. meliloti,* may be involved in organizing the root hair cytoskeleton to improve the infection thread tip growth and elongation (Pellock *et al.*, 2000, Ridge, 1993). Concurrently, several experiments with EPS⁻ and Nod⁻ strains from *R. leguminoarum* and *Sinorhizobium* species led the authors to propose that EPS are important for root tip growth, deformation of the root hair and induction of the infection thread (van Workum *et al.*, 1998, Gray *et al.*, 1991). Such functions had been assigned for many years exclusively to the Nod factors. Other authors propose that EPS also attenuate the plant defense response, allowing bacterial cells to advance towards the cortical plant cells (van Workum *et al.*, 1998, Niehaus *et al.*, 1993).

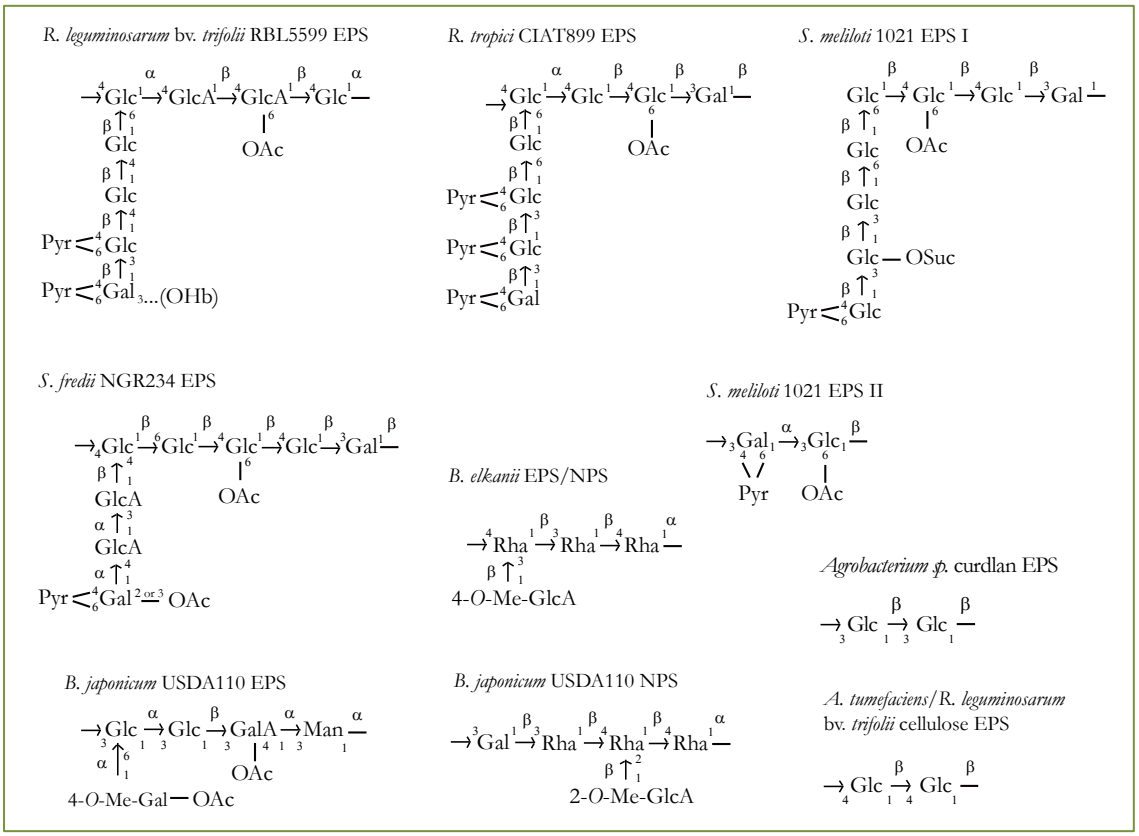


Figure 2. Chemical structure of some rhizobial EPS repeating units. Monosaccharides are abbreviated as Glc (glucose), GlcA (glucuronic acid), Gal (galactose), Rha (rhamnose). Non-carbohydrate substitutions are Ac (acetyl group), Pyr (pyruvyl group) and Suc (succynil group). Figure modified from Janczarek (2011).

In contrast to these findings, legumes that produce determinate nodules are not affected when inoculated with EPS deficient strains (Diebold & Noel, 1989, Kim *et al.*, 1989). Some authors argue that in determinate nodules, Rhizobia produce broader infection threads and only reach the outer cortical cells, whereas in indeterminate nodules rhizobia have to overcome and additional difficulty, which consist in crossing the outer

cortical layer until they reach the inner one where they infect the new cells produced by the meristem (Fraysse *et al.*, 2003, Leigh & Coplin, 1992). Bacterial EPS could be necessary for this last step. These observations should not be taken as a rule, as demonstrates the case of *S. fredii* HH103 EPS, which is not required to successfully nodulate the indeterminate-nodule forming legume *Glycyrrhiza uralensis* (Margaret-Oliver *et al.*, 2012).

Studies with different EPS mutant strains from *S. fredii* NGR234 and *R. leguminosarum* bv. *viciae* showed that the nodule development could be restored by adding exogenous EPS from other species or biovars that produce indeterminate nodules, like succinoglycan from *S. meliloti* or EPS from *R. etli* CE3, respectively. Interestingly, these substitutions were only successful when the exogenous EPS had a very similar molecular structure to the one from the native strain, indicating the importance of some structural elements for the biological activity (Gray *et al.*, 1991, Djordjevic *et al.*, 1987).

There is less knowledge regarding the role of EPS during the free-living phase of rhizobia, but it seems clear their relevance for biofilm formation and the attachment to either inert (sand, glass, plastic, quarz) or biological surfaces (Rinaudi & Giordano, 2010, Fujishige *et al.*, 2006). This has been deeply studied in *R. leguminosarum,* which is able to produce at least four different exopolysaccharides: acidic EPS, cellulose fibrils, glucomannan, and a galactose-rich gel forming EPS (Williams *et al.*, 2008). Cellulose fibrils have proven to be necessary for firm adhesion and aggregation to the tips of elongated root hairs, but they are not necessary for biofilm formation on abiotic surfaces (Laus *et al.*, 2005). Glucomannan or UPP (unipolar polysaccharide), is composed mainly by mannose and glucose, and binds specifically to lectins from pea and vetch. As the same name indicates, this EPS is only synthesized at one pole of the cells. Interestingly, this UPP seems to be important only under acidic conditions (pH 5.6), and it is not necessary for biofilm formation on abiotic surfaces (Laus *et al.*, 2006). In contrast, acidic EPS is involved in biofilm formation both in abiotic and root surfaces (Williams *et al.*, 2008). *A. tumefaciens* produces also cellulose fibrils and UPP: the increment in cellulose production provides a better attachment to abiotic or plant surfaces, whereas the UPP is involved in biofilm formation and is determinant for the irreversible attachment. The optimal production of UPP in *A. tumefaciens* takes place under low calcium, low phosphate, and acidic conditions (Matthysse, 2014). Biofilm formation in biotic and inert surfaces has proved to be dependent on EPS also in *B. japonicun* and *S. meliloti* (Rinaudi & Giordano, 2010). The latter will be discussed in detail in the last section of this Introduction.

Besides their biological role, rhizobial EPS have become a very useful tool in food, industry, and medicine sectors. They are used as thickening, emulsifying, chelating, stabilizing, or wound healing agents. Understanding EPS biosynthetic pathways will allow to improve its production or even to synthesize new polymers with the desired properties (Freitas *et al.*, 2011, Becker, 2015).

Cyclic dimeric (3'→5') GMP.

Cyclic dimeric (3'→5') GMP (c-di-GMP) is an ubiquitous bacterial second messenger that has been involved in a large number of different physiological processes, like biofilm formation, exopolysaccharide production, motility, cell cycle and virulence. Changes in c-di-GMP concentrations promote a wide range of responses comprising from transcriptional regulation of targeted genes to enzyme activity modulation (Romling *et al.*, 2013). C-di-GMP was first described as an allosteric activator of cellulose synthesis in *Gluconacetobacter xylinus* (Ross *et al.*, 1987). The description led to the identification of the enzymes responsible for the synthesis and degradation of the molecule, diguanylate cyclases (DGC) and phosphodiesterases (PDE) respectively (Tal *et al.*, 1998). DGC present a characteristic consensus domain GG(D/E)EF while PDE may present either an EAL or an HD-GYP domain. Soon after their description, genome analysis revealed that these proteins are widespread in eubacteria and, recently, a DGC has been described in eukaryotes (Romling *et al.*, 2013).

C-di-GMP synthesis and degradation.

C-di-GMP is synthesized from two GTP molecules (Figure 3). Generally, active DGCs work together as homodimers, carrying each protein a GTP joined to their GGDEF motif (the active-site). This dimer needs Mg^{+} or Mn^{+} cations to catalyze the phosphoester bond between GTP molecules. The first two glycines residues from the GG(D/E)EF are involved in GTP binding; the third residue (aspartic or glutamic acid) is essential for the catalysis and also participates in metal coordination together with the fourth glutamic residue. The own c-di-GMP regulates the activity of the DGC by binding allosterically to the inhibition site (I-site), an RXXD motif located five amino acids upstream of the active site (Chan *et al.*, 2004). Thanks to this feedback inhibition excessive consumption of GTP

is avoided and only the appropriate levels of c-di-GMP are produced. However, not all DGCs display an I-site, and little is known about how these proteins are regulated. Some cases of competitive inhibition by substrate binding to the active site have been described (Yang *et al.*, 2011).

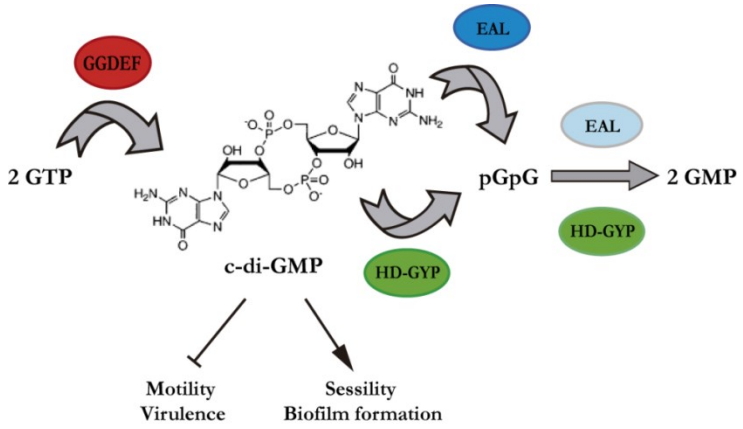


Figure 3. C-di-GMP synthesis and degradation by DGCs and PDEs. EAL phosphodiesterases convert pGpG to GMP at a slower rate than the HD-GYP ones. Generally, low c-di-GMP levels promote motility and production of virulence factors, whereas high levels induce sessility and biofilm formation. Modified from Caly *et al.* (2015)

c-di-GMP can be degraded by PDEs with either EAL or HD-GYP domains (Figure 3). For the EAL proteins, the degradation occurs in two sequential steps: first, they hydrolyze the c-di-GMP into linear di-GMP (5'-pGpG), a reaction that requires also Mg^{2+} or Mn^{2+} to catalyze the activity and is strongly inhibited by Ca^{2+}. EAL proteins can continue the degradation to GMP, but at much slower rates than the first reaction. Thus, it is suggested than *in vivo* hydrolysis of 5'-pGpG might be carried out by alternative enzymes (Caly *et al.*, 2015). Unlike DGCs, EAL proteins can be active as monomers, although the majority of proteins studied so far tend to form dimers or oligomers *in vitro*, suggesting that EAL dimers might be the most stable functional units (Romling *et al.*, 2013). Much less is known about the HD-GYP domain, since many attempts to obtain the crystal structure have failed. Just very recently, the very first crystal structure of an active HD-GYP protein has been obtained from *Persephonella marina* (Wigren *et al.*, 2014). This protein contains three iron ions in the active site which are necessary to catalyze the reaction. In contrast to the EAL domain, HD-GYP PDEs seem to catalyze the hydrolisis of c-di-GMP into GMP at a high rate, producing 5'-pGpG only transiently. The HD-GYP phosphodiesterases are not related to those who have EAL domains; they form a subfamily within the HD domain superfamily of phosphohydrolases metallo-dependent (Wigren et al., 2014).

Multidomain architectures in DGCs and PDEs.

It is not rare to find proteins that display both GGDEF and EAL domains, which hinders the prediction of their function as DGC or PDE. In some of those cases it is possible that the protein acts as a bifunctional enzyme, with one activity prevailing over the other depending either on the environmental/intracellular signals or on the intervention of accessory proteins. In others, one domain is inactive but still is able to bind the substrate (GTP) or the product (c-di-GMP). This type of "degenerate" domains can modulate the activity of the other one, or participate in protein-protein interactions or protein-ARN interactions. All this features make difficult to unravel the signaling pathways that are regulated by c-di-GMP (Romling *et al.*, 2013, Krasteva *et al.*, 2012).

Additionally, DGCs and PDEs present accessory sensory domains which may regulate their activities. The most common architecture domains include REC, GAF or PAS motifs associated with GGDEF or EAL ones. GAF and PAS domains are known to bind small ligands like chromophores, heme, flavin mononucleotide, etc., and allow the protein to sense O_2, NO, CO, light, quorum sensing molecules, or the redox state of the electron chain components, among others. On the other hand, REC domains are frequently the response regulators of two component signal transduction systems. These proteins can be cytosolic or membrane bounded, suggesting that numerous environmental and cellular signals might be integrated into c-di-GMP signaling pathways (Romling *et al.*, 2013).

C-di-GMP signaling.

As mentioned before, c-di-GMP has proved to be widespread among bacteria. However, it is remarkable the great variability regarding the number of proteins involved in the metabolism of this second messenger. Figure 4 shows a general view of the number of DGCs and/or PDEs coded in the genomes of representative species from the five major phyla. Some species like *Staphylococcus aureus* and *Haemophylus influenzae* do not present any gene related to c-di-GMP metabolism, whereas other species like *Kineococcus radiotolerans* and *Shewanella amazonensis* display up to 83 and 73 GGDEF, EAL, and HD-GYP domain-encoding genes respectively (Liang, 2015). The presence of c-di-GMP metabolism enzymes in ancient phylogenic groups such as *Cyanobacteria*, *Thermotogae* or *Chloroflexi* supports the hypothesis that c-di-GMP was adopted as a second messenger early in the evolutionary history of bacteria. Generally, bacteria that have evolved as obligate or temporal parasites

present a reduced number of DGCs or PDEs compared to those free-living ones that have to adapt constantly to environmental changes (Romling *et al.*, 2013).

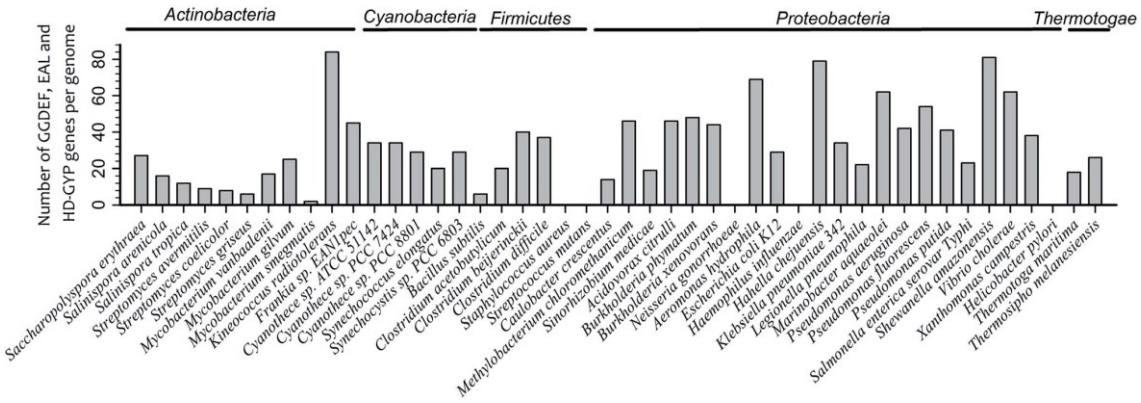


Figure 4. Number of proteins with GGDEF, EAL, and HD-GYP domains encoded among the genomes of bacteria belonging to the most representative phyla. Figure reproduced from Liang (2015).

Based on the many evidences about c-di-GMP roles compiled over the last 28 years, a general rule has been stablished. High levels of c-di-GMP promote the synthesis of polysaccharides, fimbriae, adhesins, biofilm formation, cell differentiation, and more specifically, heterocyst formation in cyanobacteria, or the survival and transmission of some intracellular pathogens. On the other hand, low c-di-GMP levels induce many motility-like processes (swimming, swarming, twitching, gliding), the production of virulence factors, toxins, expression of type three secretion systems, etc. Thus, c-di-GMP leads the transition between sessile and motile lifestyles. This transition must be very fine regulated to allow bacteria survival and adaptation to the environmental changes (Sondermann *et al.*, 2012, Hengge, 2009, Romling *et al.*, 2013). But, how is this single molecule able to regulate such a wide range of physiological processes? First, c-di-GMP is able to regulate at the transcriptional, posttranscriptional, and posttranslational levels. As outlined before, the first c-di-GMP effector described was the PilZ domain of a cellulose synthase from *G. xylinus*. Since then, others enzymes allosterically regulated by c-di-GMP have been described, but also transcription factors, flagella basal bodies, and ARNs (riboswitches)(Ryan *et al.*, 2012). Recently Trampari *et al.* (2015) demonstrated the allosteric regulation of flagellar protein export ATPases and type III/VI secretion systems. An increasing number of proteins with degenerate GGDEF or EAL domains have also been described as c-di-GMP effectors that can modulate the response of other domains located

in the same protein. Furthermore, it seems that c-di-GMP effectors present different affinity levels for the molecule. This would provide a kind of filter or threshold when it comes to generate specific responses.

The temporary and local activity of DGCs and PDEs can contribute as well to generate specific regulatory networks. Not all these proteins are present at the same time or in the same place in the cell. Their expression and activity would depend on different stimuli, and if their presence is restricted to specific sites, they would generate "local" c-di-GMP pools that trigger certain responses but not others (Hengge, 2009). In relation to this theory, different c-di-GMP signaling models have been proposed (Figure 5). If the effector is closely associated with the DGCs/PDEs, the c-di-GMP metabolism will likely be low (Figure 5A). In contrast, if the effector is distant, higher c-di-GMP levels will be required to transduce the signal (Figure 5B). Furthermore, the DGC protein can form part of the same transmembrane sensory protein, or either it can be an intermediate between the sensory protein and the target one (Figure 5C). Nevertheless, the number of identified c-di-GMP effectors to date is quite low, if we take into account the huge amount of c-di-GMP metabolizing enzymes. Finding new effectors is one of the major challenges actually, as well as characterizing the concrete mechanism by which the spatial and temporal separation of the signaling pathways is achieved.

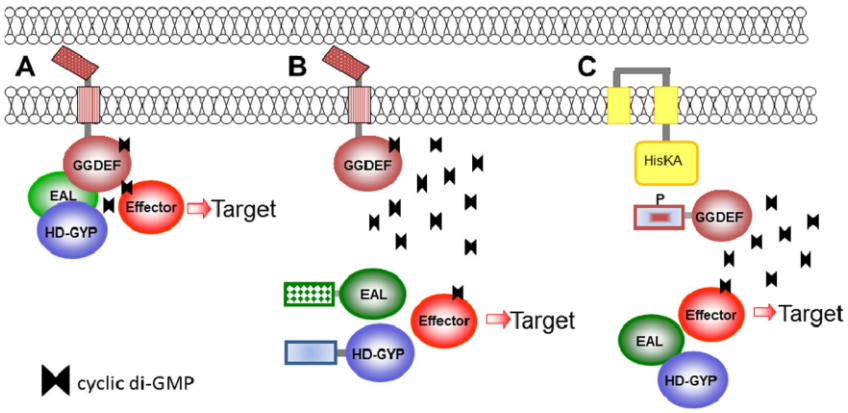


Figure 5. Different c-di-GMP signalling models. Proteins involved in c-di-GMP metabolism can form complexes very close to the target (A), or they can act from the distance (B). In the first case local low c-di-GMP levels are often sufficient, whereas in the second case the production must be higher to assure the signal transduction. In a third scenario, DGCs act as intermediary betwee the sensor membrane protein and the target. Figure reproduced from (Romling *et al.*, 2013).

Quorum sensing in bacteria.

Bacteria are able to sense the presence of their relatives and produce coordinated responses when the population reaches certain growth stages or environmental conditions. This process is known as quorum sensing (QS), and relies in the production of diffusible signal molecules called autoinducers. As the population density increases, so does the synthesis of signal molecules. They accumulate in the extracellular surroundings and eventually enter inside the cells by diffusion. Once autoinducers attain a threshold accumulation, they trigger a signal cascade that ultimately affects the expression/repression of concrete genes (Williams, 2007, Fuqua *et al.*, 1994). The best characterized QS systems are found in gram-negative bacteria, which use acylated homoserine lactones (AHL) as signal molecules or autoinducers. Other molecules such as furanones, fatty acid methyl esters or long chain fatty acids have also been described as bacterial communicators (Williams *et al.*, 2007)

The amphipathic nature of AHLs molecules allow them to cross the phospholipid bilayer of the cell membranes an also to freely move along the intracellular and extracellular environments. AHLs can vary depending on the length of the acyl chain (from 4 to 18 carbons) or the substitutions of the carbon chain. One organism can produce different AHLs, but also different species can produce the same AHL. Thus, quorum sensing cross-talking between different species might take place, affecting the behaviour of one or both populations. For instance, the production of the antibiotic phenazine by *Pseudomonas aureofaciens* depends not only on its own AHLs, but it also responds to signal molecules produced by other plant-associated bacteria (Pierson *et al.*, 1998).

The first QS system was described in *Vibrio fischeri*, a bioluminescent marine bacterium that establishes symbiotic relationships with some fishes and squids (Eberhard *et al.*, 1981). All AHLs QS systems identified so far present the basic composition of the *V. fischeri* one: a LuxI-type HLA synthase and a LuxR-type regulator. There are two other families of AHL synthases (LuxM and HdtS), although the LuxI-type is the most spread. On the other hand, LuxR-type regulators present two characteristic domains: the first one recognizes and binds AHLs, and the other one is a DNA binding motif. Generally, AHL interaction provokes conformational changes in the protein that allow its binding to specific DNA regions and the transcriptional activation. AHLs synthases genes are also induced by this system, leading to a positive feedback that increment the signal molecules accumulation. In some cases, there is a histidine kinase instead of the Lux-R type regulator that initiates a

phosphorelay signal transduction cascade once it recognizes the corresponding AHL. This cascade ultimately leads to the expression/repression of concrete genes (Williams, 2007, Bassler, 1999).

QS contributes to environmental adaptation by coordinating bacteria populations in producing virulence or plant biocontrol determinants, biofilm formation, collective defence against competitors, etc. In addition it provides information about local environment or spatial distribution of the cells (Williams, 2007). Signal molecules can also influence in the behaviour of fungal, plant, or animal cells (Telford *et al.*, 1998, Hogan *et al.*, 2004, Bauer & Mathesius, 2004). Among all these traits we will focus on the production of exopolysaccharides.

Quorum Sensing, EPS, and c-di-GMP networks.

As mentioned earlier in this section, EPS are one of the major components of biofilms. Their production is generally associated with sessile stages, when bacteria tend to aggregate, i.e., to increase the cell density. Thus, QS systems play an essential role in terms of establishing the appropriate moment to initiate the production of EPS. On the other hand, c-di-GMP has been proved to regulate a wide range of different bacterial polysaccharides, as it will be discussed below. Several evidences suggest the existence of regulatory connections between these three elements (QS, c-di-GMP and EPS) which can present different cooperation and complexity levels (Srivastava & Waters, 2012).

The first direct connection between QS and c-di-GMP was described in the plant pathogen *Xanthomonas campestris*. This bacterium produces an auntoinducer known as Difusible Signal Factor (DSF) which is recognized by the sensor kinase RpfC. Only when RpfC binds DFS, it phosphorylates the REC domain from the HD-GYP protein RpfG, provoking a decrease in ci-di-GMP levels. When RpfC does not recognize DFS, it interacts with the DFS synthase (RpfF) and prevents the synthesis of DFS. Therefore, low cell densities are associated with high c-di-GMP levels and vice versa. The lowering of c-di-GMP levels triggers the production of virulence factors (among them EPS) through the transcription factor Clp. Interestingly, RpfG was the first HD-GYP protein proved to be a PDE, and it seems to directly interact with other c-di-GMP metabolizing enzymes in *X. campestris*. By this way, the initial quorum sensing signal is integrated and amplified through the c-di-GMP regulatory network (Fouhy *et al.*, 2006, Ryan & Dow, 2010).

Another example of the connection between QS and c-di-GMP is described in *Vibrio cholerae*, although in this case is more complex and occurs at the transcriptional level. Similarly to what it is described in *X. campestris*, high c-di-GMP levels are associated with low cell densities, when the exopolysaccharide VPS is produced. The transcription of 14 different GGDEF and EAL proteins and four HD-GYP domain containing proteins is controlled by HapR, a high-cell-density transcription factor. The expression of HapR is regulated by small ARNs (*qrr*) that at the same time are controlled by the phosphorylation stage of LuxO, the central response regulator of the QS system. QS and c-di-GMP also control the expression of two key transcription factors: AphA and VpsT. The promoter regions of *aphA* and *vpsT* are binding targets of the mentioned HapR but also of VpsR, a transcription activator that binds c-di-GMP. The binding to the promoter regions is mutually exclusive. Thus, QS and c-di-GMP signalling are integrated into a very fine-tuning network that is able to sense and responds to a wide range of environmental inputs (Srivastava *et al.*, 2011, Waters *et al.*, 2008).

Besides its relation with QS systems, c-di-GMP can also contribute to biofilm formation by directly regulating the transcription and/or expression of EPS related genes. Allosteric regulation through PilZ domains have been described also in the alginate synthase from *P. aeruginosa* (Oglesby *et al.*, 2008). The protein PelD, necessary for the synthesis of another EPS from *P. aeruginosa* (Pel EPS) binds c-di-GMP through a degenerate GGDEF domain. The biosynthetic operon of this EPS is also transcriptionally regulated by FleQ, a transcriptional activator which activity relays on c-di-GMP binding (Hickman & Harwood, 2008, Li *et al.*, 2012). Other examples of transcriptional regulation by c-di-GMP are curdlan synthesis in *A. tumefaciens* and poly-N-acetylglucosamine (PNAG) in *E. coli.* (Liang, 2015, Romling *et al.*, 2013).

In relation to the Rhizobia group, several publications have remarked the importance of QS systems, which could participate in preparing/coordinating bacteria during the establishment of the symbiotic relation with their host legumes. There are strong evidences suggesting that QS systems participates in nodulation efficiency, EPS production, symbiosome development, plasmid transfer, stationary-phase adaptation, nod genes regulation, or nitrogen fixation, although the concrete mechanisms of action remain unclear (Gonzalez & Marketon, 2003). Interestingly, rhizobia produce the greatest variability of AHLs described so far, and also present many different QS systems that vary

even between isolates from the same species. This makes difficult to stablish general rules about which traits are regulated by QS (Sanchez-Contreras *et al.*, 2007).

The role of c-di-GMP in Rhizobia has been less studied, however. High c-di-GMP levels are known to increase cellulose production in *R. leguminosarum* and *A. tumefaciens* (Ausmees *et al.*, 2001). On the other hand, Perez-Mendoza *et al.* (2014) found that high c-di-GMP levels provoke a reduction in swimming motility, increase in biofilm formation, and better attachment to root surfaces in *R. leguminosarum* and *R. etli.* However, this condition seemed to be deleterious for the symbiotic efficiency. Besides functional analysis, Gao *et al.* (2014) performed a genome analysis of six representative rhizobia strains and found a high number of proteins involved in c-di-GMP synthesis and degradation, suggesting the relevance of this second messenger in rhizobial signal transduction.

Sinorhizobium meliloti.

In the thesis project presented below, we have used *Sinorhizobium meliloti* as organism of study. As outlined before, this α-proteobacteria belongs to the Rhizobia group. Its genome comprises three replicons: one chromosome of 3.65 Mb and two megaplasmids called pSymA (1.35 Mb) and pSymB (1.68 Mb) (Galibert *et al.*, 2001). Most of the genes involved in the nodulation process are located in the pSymA plasmid (Barnett *et al.*, 2001), whereas pSymB harbors a remarkable amount of genes associated to the synthesis of polysaccharides (Finan *et al.*, 2001). *S. meliloti* lives in the soil and is able to establish symbiotic relationships with various species of legumes from the genera *Medicago*, *Melilotus* and *Trigonella* (Hirsch *et al.*, 2001), being the model *S.meliloti-Medicago* one of the most studied (Jones *et al.*, 2007) .

S. meliloti is able to produce two symbiotically active exopolysaccharides, succinoglycan (EPS I) and galactoglucan (EPS II). During the last three decades a great effort has been made to characterize biochemically and genetically the synthesis of these exopolysaccharides and also their physiological roles that they play in symbiosis and in free living style. EPS I consists of a polymer of an octasaccharide repeating unit, formed by seven glucoses and one galactose decorated with pyruvil, acetyl and succinyl modifications (figure x). EPS II is formed by a disaccharide repeating monomer, one galactose and one glucose with pyruvil and acetyl modifications (Gonzalez *et al.*, 1996b, Gibson *et al.*, 2008). Both polysaccharides can be polymerized into either high (HMW) or low molecular weight

forms (LMW). LMW exopolysaccharides are the symbiotically active forms that are needed for the first steps in the symbiosis and seem to play an important role in the bacterial interactions with the environment. (Lehman & Long, 2013, Rinaudi & Gonzalez, 2009, Battisti *et al.*, 1992, Gonzalez *et al.*, 1996a, Lloret *et al.*, 1998)

The structural genes involved in their synthesis are both located in separate regions in pSymB, the largest megaplasmid. Interestingly, around 12% of the genes annotated in the pSymB seem to be involved in the synthesis of cell surface carbohydrates, but the putative polysaccharides produced have not been characterize yet (Finan *et al.*, 2001, Galibert *et al.*, 2001). The *exp* genes responsible for the EPS II synthesis constitute a 32 kb cluster organized in four transcriptional units: *wga* (*expA*), *wgd* (*expD*), *wge* (*expE*), and *wgcA* (*expC*)-*wggR* (*expG*) (Glazebrook & Walker, 1989, Becker *et al.*, 1997). The first three operons are required for the structural biosynthesis, whereas *wgcA* and *wggR* code for a glycosyl transferase and a transcriptional regulator. EPS II production is tightly regulated by several systems. Low-phosphate conditions characteristic of soils are sensed by the PhoR/PhoB system, which triggers the transcription of EPS II genes in cooperation with the WggR transcriptional regulator. MucR, in contrast, acts as a negative regulator by binding to the promoter regions of the biosynthesis (Bahlawane *et al.*, 2008, Becker *et al.*, 2002).

The *exo*/*exs* genes cluster (≈35 kb) responsible for the synthesis of the succinoglycan is located 200 kb apart from the *exp* genes, although there are some other genes necessary for the synthesis and regulation mapped on the chromosome (*exoC*, *exoR*, *exoS*, *exoD*). Likewise the *exp* genes, it is organized in several transcriptional units that harbors all the genes necessary for the synthesis of precursors, unit assembly, non-sugar modifications, polymerization, and transport (Janczarek, 2011, Jones *et al.*, 2007). Several genes have been identified as regulators of the EPS I synthesis: *exoX*, *exsB*, *exoR*, *exoS*, *cbrA* and *emmC* act as negative regulators both at the transcriptional and post-transcriptional levels, whereas *syrM, exoD* and *Muc*R act as positive transcriptional regulators (Janczarek, 2011). The EPS I production is stimulated by high phosphate levels, hyperosmotic stress, and low availability of ammonia and sulfur among other nutrients. It is worthy to distinguish that MucR inversely regulates EPS I and EPS II depending on the environmental conditions (Bertram-Drogatz *et al.*, 1998, Keller *et al.*, 1995).

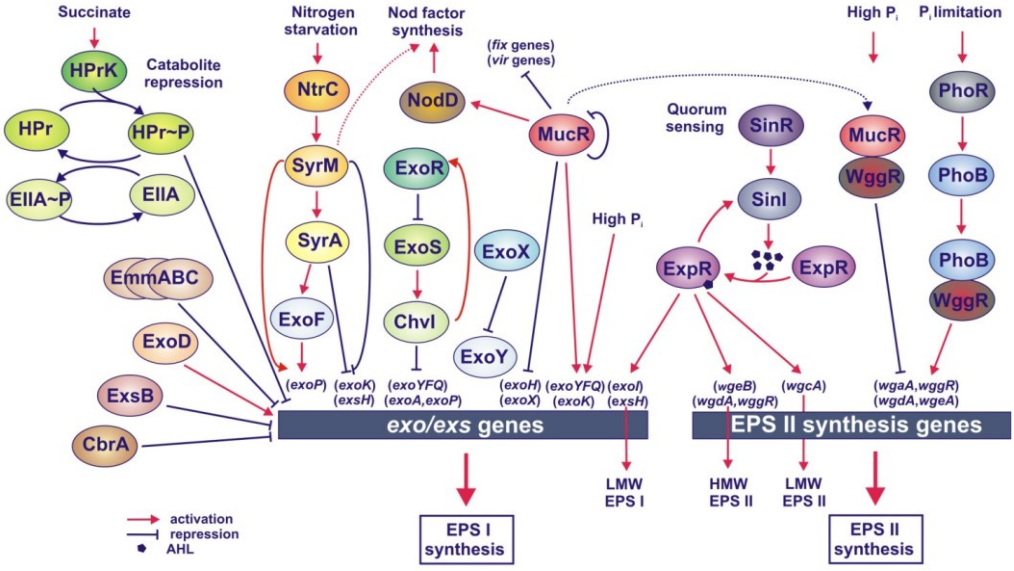


Figure 6. Model of the regulatory network that controls the production of EPS I and EPS II. Figure reproduced from (Janczarek, 2011).

Several studies pointed out the importance of the quorum system Sin/ExpR as a master regulator for the synthesis of EPS II and also regulating EPS I production (Pellock *et al.*, 2002, Glenn *et al.*, 2007). The system consists in SinI, a long chain *N*-acyl homoserine lactone synthase and two LuxR-type transcriptional regulators, SinR and ExpR (Marketon *et al.*, 2002, Hoang *et al.*, 2004). SinR, whose gene is located immediately upstream of *sinI*, acts as a typical transcriptional regulator of AHL synthase (Marketon *et al.*, 2002). *expR*, separated more than 1.5 Mb from the other two genes, is truncated in the reference strain Sm1021 and the closely related Sm2011 by an insertion sequence (Pellock *et al.*, 2002) The restoration of the truncated gene revealed the importance of this transcriptional regulator in multiple traits apart from EPS regulation like motility, nitrogen fixation or metal transport (Hoang *et al.*, 2004, Hoang *et al.*, 2008). Regarding exopolysaccharide regulation, the Sin/ExpR system induces the expression of the EPS II genes when bacteria reach a certain population density, by liberating them from the inhibition of MucR. An alteration in *expR,* as it happens with Sm1021 or Sm2011, completely abolishes the EPS II production (Pellock *et al.*, 2002). On the other hand, the regulation of EPS I by the Sin/ExpR system is not as drastic as with the EPS II. In this case the QS system seems to act as a fine-tuning modulator of the LMW fraction (Glenn *et al.*, 2007). As a summary, the production of both EPS I and EPS II is tightly regulated by a complex set of networks that include two

component systems, quorum sensing, and several transcription regulators. Some pathways overlap with each other whereas others are independent. By this way *S. meliloti* is able to respond and adapt to the very inconstant environments like the soil and the rhizosphere. Figure 6 shows a general scheme of how these two EPS are regulated

In relation to the biological role of EPS I and EPS II, each one is sufficient to mediate the infection of root hairs and the elongation of the infection thread, although succinoglycan has been proved to be more efficient (Pellock *et al.*, 2000). As mentioned before, mutants defective in the production of these polysaccharides are not able to reach the nodule, which yields a Fix^- phenotype. In both exopolysaccharides the LMW form is the symbiotic active form (Battisti *et al.*, 1992, Gonzalez *et al.*, 1996a). Regarding the free-living stage, EPS II has important roles in attachment to inert or biotic surfaces, aggregation of planktonic cells, root colonization and biofilm formation. Again, the LMW form is essential for these processes (Sorroche *et al.*, 2010, Rinaudi & Gonzalez, 2009) . Recently, Dilanji *et al.* (2014) described a new type of "passive" motility driven by the HMW of the EPS II. The authors explain that characteristic patterns of Sm8530 spreading colonies in semi-solid medium are produced by two physical chemical processes: osmotic pressure and depletion forces. By this way cells are able to ride passively with the mucoid layer as it moves downwards without spending energy. They also suggest that this phenomenon might occur in the natural habitat of the bacteria, giving the HMW form of the EPS II a biological role that has been elusive. Finally, there are evidences showing that EPS I and EPS II can also protect the cells against H_2O_2 (Lehman & Long, 2013).

Recently it has been shown that *S. meliloti* produces a third exopolysaccharide whose synthesis is also regulated by the Sin/ExpR system (Perez-Mendoza *et al.*, 2015). The polysaccharide consists of a linear mixed-linkage (1→3) (1→4)-β-D-glucan (ML β-glucan). This structure is similar to other β-glucans found in some lichens (lichenan) and cereals (barley glucan), but it is the first report among bacteria. In addition, the ML β-glucan presents a unique primary structure. Plant ML β-glucans normally display certain ratios of cello-oligosaccharides with different degrees of polymerization, linked by (1→3) bonds, whereas the ML β-glucan from *S. meliloti* only contains a →3)-β-D-Glc*p*-(1→4)-β-D-Glc*p*-(1→ repeating unit. Two genes located in a single operon in the pSymB plasmid have been involved in its synthesis, *bgsB* and *bsgA*. BgsB contains a HlyD-like domain and might function in the export of the polysaccharide, whereas BgsA is annotated as a cellulose synthase and is likely to be the glucosyltransferase that synthesizes the ML β-glucan. The

operon is conserved among bacteria from genera *Methylobacterium*, *Agrobacterium*, *Rhizobium* and *Sinorhizobium*. The production of the ML β-glucan is boosted when the levels of cyclic diguanylate (c-di-GMP) are artificially increased. Under these conditions, the ML β-glucan provides a better attachment to the root surfaces (Perez-Mendoza *et al*., 2015).

OBJECTIVES

During the last two decades, c-di-GMP has gone from being just an activator of the synthesis of cellulose to an important ubiquitous bacterial second messenger that regulates important traits in the lifestyles of these microorganisms. At the beginning of this project, however, there was no information about the role that c-di-GMP had among free living and symbiotic rhizobia.

The first aim of this work was to study the role that proteins involved in c-di-GMP metabolism play in *S. meliloti* lifestyles, regarding exopolysaccharide production, biofilm formation, motility and symbiotic properties. Once the first results were obtained, we focused on the characterization of SMb20447, a putative diguanylate cyclase regulating the synthesis of a new expolysaccharide. The specific goals of this project are:

1. Characterization of the exopolysaccharide affected by the activity of SMb20447.
2. Identification of other genes implicated in the synthesis of the exopolysaccharide.
3. Unravelling the signalling network by which c-di-GMP regulates the exopolysaccharide production.

MATERIALS & METHODS

Growth conditions, antibiotics, and dyes.

Regular growth of *Escherechia coli* strains was carried out in LB broth (10 g of tryptone, 5 g of yeast extract, and 5 g of NaCl per liter) at 37 ℃ , whereas *Sinorhizobium meliloti* strains were gown in TY broth (5 g of tryptone, 3 g of yeast extract, and 0.4 g of $CaCl_2$ per liter) at 28℃. Solid TY or LB for plates was prepared by adding 15 g/liter of American bacteriological agar prior to sterilisation. Other media used for specific purposes are listed in Table 1.

Table 1. Media used for specific purposes.

Minimal Medium

K_2HPO_4	0.3 g
KH_2PO_4	0.3 g
$MgSO_4*7H_2O$	0.15 g
$CaCl_2*2H_2O$	0.05 g
$FeCl_3$	0.006 g
NaCl	0.05 g
Sodium glutamate	1.1 g
Mannitol	10 g
Distilled water	1000 ml

Adjust pH to 6.8-7.2
Sterilize 20 min/121°C
Add 1ml of vitamin stock

Vitamins stock solution (1000 x)

Biotin	0.2g
Thiamine hydrochloride	0.1 g
Sodium pantothenate	0.1 g
Distilled water	1000ml

Sterilize by filtration

BNM (Buffered Nod Medium)

$CaSO_4*2H_2O$	344 mg
MES	390 mg
Nod Majors (200 x stock)	5 ml
Nod Minors I (200 x stock)	5 ml
Nod Minors II (200 x stock)	5 ml
Fe-EDTA (200 x stock)	5 ml
Distilled water	980 ml

Adjust pH to 6.5
Add 11.4 of purified agar
Sterilize 30 min/ 121°C

200 x stock solutions (per L)

Nod Majors

$MgSO_4*7H2O$	24.4 g
KH_2PO_4	13.6 g

Nod Minors I

$ZnSO_4*7H_2O$	0.92 g
H_3BO_3	0.62 g
$MnSO_4*H_2O$	1.69 g

Nod Minors II

$Na_2MoO_4*2H_2O$	50 mg
$CuSO_4$	3.2 mg
$CoCl_2*6H_2O$	5 mg

Fe-EDTA

Na_2EDTA	3.73 g
$FeSO_4*7H_2O$	2.78 g

Sterilize for 30 min at 121°C

YMB (Yeast Mannitol Broth)

K_2HPO_4	0.5 g
$MgSO_4*7H_2O$	0.2 g
NaCl	0.1 g
Yeast extract	0.4 g
Mannitol	10 g
Distilled water	1000ml

Adjust pH to 6.8-7
Sterilize 30 min/121°C

Solid YMB

Add 15 g purified agar prior sterilization

When required, antibiotics and other compounds were added at the following final concentrations: streptomycin (Sm), 300 μg/ml; gentamicin (Gm), 10 μg/ml for *E. coli* and 25-50 μg/ml for *S. meliloti*; kanamycin (Kn), 25 μg/ml; neomycin (Nm), 100 μg/ml;

tetracycline (Tc), 10 μg/ml; ampicillin (Amp), 100 μg/ml; chloramphenicol (Cm), 30 μg/ml; thrimetroprim (Tmp), 100 μg/ml for *E. coli* and 200-400 μg/ml for *S. meliloti*; calcofluor (CF), 200 μg/ml; congo red (CR), 50 μg/ml; aniline blue, 50 μg/ml; X-gal, 40 μg/ml; cumate, 10-50 μM; arabinose, 2%.

Molecular techniques.

DNA manipulation.

Genomic DNA and plasmid extractions were carried out with the corresponding kits attending to the manufacturer's instructions (REALPURE® from Durviz S.L. and Wizard® *Plus* SV minipreps from Promega Corp.), as well as all the enzymatic digestions and cloning steps (Takara Bio Company© and Promega Corp.). Plasmid minipreps were also performed according to Engebrecht *et al.* (2001). For regular PCR reactions, a polymerase from Biotools (B & M Labs, S.A.) was used, whereas a more reliable one was chosen for amplification of fragments used in complementation tests, overexpressions, generation of mutants or protein assays (Hot Start High Fidelity PCR System from Roche Diagnostics). Generally, PCR conditions consisted of one initial denaturing step at 95ºC for 3 min, 25-30 cycles of denaturing (95ºC/30s), annealing (optimal oligo temperature/30s), and elongation (1 min/kb at 72ºC), and a final elongation step at 72ºC for 7 min. All primers used are listed in Table 4. Generally, all primers worked well using annealing temperatures between 59°C and 60° C. PCR or digestion products were subjected to electrophoresis in 0.8-1% agarose gels, and, if needed, they were purified using the Wizard® SV Gel and PCR Clean-Up System from Promega following the manufacturer's indications.

Construction of *S. meliloti* strains and plasmids.

Plasmids and strains used in this project are listed in Tables 2 and 3, respectively. Mutations from *S. meliloti* 2011 were transduced into *S. meliloti* 1021 or 8530 using the bacteriophage ϕM12 (laboratory collection) as described by Finan *et al* (1984).

Regarding the insertion mutants, an internal fragment (around 500 bp) of the desired gene was amplified, cloned into pGEM®-T Easy (Promega) and subcloned into pK18*mob*. The resulting derivatives were stored in *E. coli* DH5α and transferred to *S. meliloti* by

triparental mating (Glazebrook & Walker, 1991) using another DH5α strain harbouring plasmid pRK600 as a helper. Neomycin-resistant transconjugants, where a single homologous recombination had taken place, were selected and verified by PCR.

Deletion and *lacZ* fusion mutants were obtained by double homologous recombination events. Briefly, around 1kb up and down-stream fragments of the desired genes were amplified, cloned into pGEM®-T Easy and subcloned into pK19*mobSacB* to generate in-frame deletion constructs. When required, the lacZ-Gn cassette was excised from plasmid pAB2001 and ligated to the corresponding constructs between the up and down-stream fragments (or in some cases in the middle of a fragment containing the entire gene). Again, plasmid derivatives were introduced into *S. meliloti* by triparental mating. Neomycin-resistant transconjugants were gown in liquid TY medium without antibiotics for 24 hours and several dilutions were plated on TY solid medium with 10% sucrose. (Schafer *et al.*, 1994). Neomycin-sensitive colonies were checked by PCR and Southern blot (Brown, 2001).

For the complementation and overexpression assays, a complete copy of the desired gene was amplified and cloned into pGEM®-T Easy. Several clones were sequenced to verify the fidelity of the sequence. Then, the insert was subcloned into pFAJ1708 or pBBRMCS-5 and introduced in the desired strains by triparental mating. Transconjugants were selected on TY plates with tetracycline or gentamicin.

General mutagenesis.

The screening of genes implicated in the synthesis or regulation of the ML β-glucan in Sm8530 was tackled by general mutagenesis, using a Tn*5*-Kanamycin transposon from plasmid pSUP2021 (Simon *et al.*, 1983). The mutagenesis was grown in liquid TY medium with Km and plated on minimal medium (Robertsen) supplemented with Congo red (125 μg/ml). Red colonies were selected and re-streaked several times in solid medium with Congo Red. Kanamycin resistance of these colonies was also checked. Candidates were kept and subjected to genomic DNA extraction followed by digestion with the restriction enzyme StuI. This enzyme produces a 12Kb fragment that contains the entire *SMb20447-SMb20452* region. A Southern blot was performed with a probe that binds between genes *SMb20450* and *SMb20451* (806 bp DNA fragment amplified with primers OB87 and OB88).

Table 2. Plasmids used in this project.

Plasmid	Relevant characteristics	Source
pRK600	Helper plasmid for triparental mating; Cm^r	Finan *et al.* (1986)
pGem T-easy	Cloning vector; Amp^r	Promega Corp.
pK18*mob*	Suicide vector for insertion mutations; Kn^r	Schafer *et al.* (1994)
pK19*mobsacB*	Suicide vector for allelic exchange; Kn^r	Schafer *et al.* (1994)
pBBRMCS-5	Replicative plasmid with *lacZ* promoter for overexpression; Gn^r	Kovach *et al.* (1995)
pSUP2021	Vector containing a Tn5-Nm transposon for general mutagenesis; Amp^r	Simon *et al.* (1983)
pAB2001	Vector containing a gentamicin-resistance and *lacZ* promoter-probe cassette for insertion mutagenesis and generation of transcriptional fusions.	Beker *et al* (1995)
pFAJ1708	Broad-host-range plasmid with *nptII* promoter for overexpression; Tc^r	Dombrecht *et al.* (2001)
pMLBAD	Expression vector inducible with arabinose; Tmp^r	Lefebre and Valvano (2002)
pQF	Expression vector inducible with cumate; Tc^r	Kaczmarczyk *et al.* (2013)
pLS-257-1	pQF derivative containing a StrepII tag for cloning NdeI-NsiI fragments and generate translational fusions at the C-terminus of the protein; Tc^r	Claude Bruand collection, INRA
pLS-258-1	pQF derivative containing a c-Myc tag for cloning NdeI-NsiI fragments and generate translational fusions at the C-terminus of the protein; Tc^r	Claude Bruand collection, INRA
pLS-259-1	pQF derivative containing a HA tag for cloning NdeI-NsiI fragments and generate translational fusions at the C-terminus of the protein; Tc^r	Claude Bruand collection, INRA
pJL2101	pK19*mobSacB* derivative bearing an EcoRI-BglII-EcoRI in frame deletion fragment of *SMb20447*; Kn^r	This study
pJL2102	pK19*mobSacB* derivative bearing an EcoRI-NotI-EcoRI in frame deletion fragment of *SMb20448*; Kn^r	This study
pJL2103	pK19*mobSacB* derivative bearing an SphI-BamHI-SphI in frame deletion fragment of *SMb20449*; Kn^r	This study
pJL2104	pK19*mobSacB* derivative bearing an BamHI-SphI-HindIII in frame deletion fragment of *SMb20450*; Kn^r	This study
pJL2105	pK19*mobSacB* derivative bearing an SalI-SmaI-EcoRI in frame deletion fragment of *SMb20451*; Kn^r	This study
pJL2106	pK19*mobSacB* derivative bearing an PstI-HindIII-EcoRI in frame deletion fragment of *SMb20452*; Kn^r	This study
pJL2107	pK19*mobSacB* derivative bearing an BamHI-SphI-SalI in frame deletion fragment from *SMb20450* to *SMb20452*; Kn^r	This study
pJL2108	pBBRMCS-5 with *SMb20447* cloned at BamHI-KpnI sites; Gn^r	This study
pJL2108E215A	pJL108 derivative with a point mutation that substitutes the glutamic acid of the AGDEF domain of SMb20447 with an alanine; Gn^r	This study
pJL2109	pBBRMCS-5 with *SMb20448* cloned at EcoRI site; Gn^r	This study
pJL2110	pBBRMCS-5 with *SMb20449* cloned at EcoRI site; Gn^r	This study
pJL2111	pFAJ1708 with *SMb20450* cloned at the BamHI site; Tc^r	This study
pJL2112	pBBRMCS-5 with *SMb20451* cloned at SacI site; Gn^r	This study
pJL2113	pBBRMCS-5 with *SMb20452* cloned at EcoRI site; Gn^r	This study
pJL2113S61A	pJL2113 derivative with a point mutation that substitutes the serine residue at position 61 with an alanine; Gn^r	This study
pJL2114	pFAJ1708 with *SMb20391* cloned at BamHI-PstI sites; Tc^r	This study

Plasmid	Relevant Characteristics	Source
pJBPleD*	pJB3Tc19 derivate bearing a 1423 bp XbaI-EcoRI fragment containing *pleD**; Ap^r, Tc^r	Perez-Mendoza *et al.* (2014)
pJL2115	pK18*mob* derivative with a 579 bp EcoRI fragment of *SMb20447*; Kn^r	This study
pJL2116	pK18*mob* derivative with a 480 bp EcoRI fragment of *SMb20450*; Kn^r	This study
pJL2117	pK18*mob* derivative with a 410 bp EcoRI fragment of *SMb20460*; Kn^r	This study
pJL2118	pK19*mobSacB* derivative bearing a PstI fragment containing *SMb20447* interrupted with a *lacZ*-Gn cassette at the EcoRV site; Kn^r	This study
pJL2119	pJL2104 derivative with a *lacZ*-Gn cassette inserted at the SphI site; Kn^r	This study
pJL2120	pK19*mobSacB* derivative bearing an EcoRI fragment of *SMb20391* with a *lacZ*-Gn cassette inserted at KpnI site; Kn^r	This study
pJL2121	pMLBAD derivative expressing a StrepII tagged version of SMb20447 at the N-terminus; Tmp^r	This study
pJL2122	pMLBAD derivative expressing an HA tagged version of SMb20451 at the N-terminus; Tmp^r	This study
pJl2123	pMLBAD derivative expressing a c-Myc tagged version of SMb20452 at the N-terminus; Tmp^r	This study
pJL2124	pLS-257-1 derivative expressing *SMb20447*.	This study
pJL2125	pLS-257-1 derivative expressing *SMb20451*.	This study
pJL2126	pLS-257-1 derivative expressing *SMb20452*.	This study
pJL2127	pLS-258-1 derivative expressing *SMb20447*.	This study
pJL2128	pLS-258-1 derivative expressing *SMb20451*.	This study
pJL2129	pLS-258-1 derivative expressing *SMb20452*.	This study
pJL2130	pLS-259-1 derivative expressing *SMb20451*.	This study
pJL2131	pLS-259-1 derivative expressing *SMb20452*.	This study

Site-directed mutagenesis.

Point mutation at the glutamic acid residue in the AGDEF domain of SMb20447 was achieved by two rounds of PCR reactions. Firstly, primers OB129 and OB130 containing the desired change were used with external primers OB74 and OB73 respectively. The resulting PCR products were used as a single template in a second PCR round using only the external primers OB73 and OB74 that amplify the entire gene. The corresponding band was checked in an agarose gel, purified, and cloned into pGEM®-T Easy. Several clones were sequencing and one good candidate was selected for subcloning into pBBRMCS-5.

Substitution of the serine at 61 position from SMb20452 with an alanine was performed according to the instructions from the Q5® Site-Directed Mutagenesis Kit (New England Biolabs Inc.), and using primers with the desired nucleotide changes (Table **4**). The PCR product was checked in an agarose gel, purified, and cloned into pGEM®-T Easy. Several clones were sequencing and one good candidate was selected for subcloning into pBBRMCS-5.

Construction of tagged proteins.

Forward primers containing a KpnI site followed by an ATG start codon and the corresponding tag sequence (StrepII, HA or c-Myc) were used in combination with reverse primers containing HindIII sites (Table **4**). PCR products were cloned into pGEM®-T Easy and the sequences of several clones were verified. A selection of good clones and the expression vector pMLBAD were digested with the KpnI isoschizomer Acc65 and the cohesive ends were refilled using the Klenow fragment from the *E. coli* DNA polymerase I (Thermo Scientific Inc). Finally, samples were digested with HindIII to generate blunt-HindIII fragments. The purpose of this strategy was to obtain in-frame translational fusions with the Ribosome Binding Site located in the pMLBAD vector. The sequences of the resulting pMLBAD derivatives were verified.

Tagged proteins were also obtained by cloning the desired genes as NdeI-NsiI fragments into three modified versions of plasmid pQF (Laurent Sauviac, INRA-Toulouse). Each version contains one of the tag above mentioned to generate translational fusions at the C-terminus of the protein. The sequences of all constructs were verified.

Table 3. Strains used in this project.

Strains	Relevant characteristics	Source
S. meliloti strains		
Sm1021	SU47; Sm^t	Laboratory collection
Sm8530	Sm1021 *expR$^+$;* Sm^r	Juan González
Rm10002	Sm8530 *ngaB::Tn5*; Nm^r	Juan González
Rm11603	Sm8530 *exoY210::Tn5-233*	Juan González
Rm11605	Sm8530 *exoY210::Tn5-233, ngaB::Tn5;* Sm^r,Nm^r,Gm^r	Juan González
IBR1100	Sm1021 *exoY::Tn5*; Sm^r,Nm^r	This study
IBR1101	Sm8530 *SMb20447::Tn5*; Nm^r	This study
IBR1102	Sm8530 *SMb20450::pk18mob*	This study
IBR1103	Sm8530 Δ*SMb20450, SMb20460::pk18mob;* Sm^r	This study
IBR500	Sm8530 Δ*SMb20447;* Sm^r	This study
IBR501	Sm8530 Δ*SMb20448;* Sm^r	This study
IBR502	Sm8530 Δ*SMb20449;* Sm^r	This study
IBR503	Sm8530 Δ*SMb20450;* Sm^r	This study
IBR504	Sm8530 Δ*SMb20451;* Sm^r	This study
IBR505	Sm8530 Δ*SMb20452;* Sm^r	This study
IBR506	Sm8530 Δ*SMb20450, exoY210::Tn5-233;* Sm^r, Gm^r	This study
IBR507	Sm8530 Δ*SMb20450, ngaB::Tn5;* Sm^r, Nm^r	This study
IBR508	Sm8530 Δ*SMb20450 , exoY210::Tn5-233, ngaB::Tn5;* Sm^r,Nm^r,Gm^r	This study
IBR509	Sm8530 *SMb20450::lacZ-Gn;* Sm^r, Gn^r	This study
IBR510	Sm1021 *SMb20450::lacZ-Gn;* Sm^r, Gn^r	This study
IBR511	Sm8530 *SMb20447::lacZ-Gn;* Sm^r, Gn^r	This study
IBR512	Sm1021 *SMb20447::lacZ-Gn;* Sm^r, Gn^r	This study
IBR513	Sm8530 *SMb20391::lacZ-Gn;* Sm^r, Gn^r	This study
IBR514	Sm1021 *SMb20391::lacZ-Gn;* Sm^r, Gn^r	This study
IBR515	Sm8530 Δ*SMb20450, SMb20391::lacZ-Gn;* Sm^r, Gn^r	This study
IBR516	Sm8530 Δ*SMb20450, SMb20447::pK18mob, SMb20391::lacZ-Gn;* Sm^r, Gn^r, Nm^r.	This study
IBR517	Sm8530 Δ*SMb20450, SMb20447::lacZ-Gn;* Sm^r, Gn^r	This study
IBR518	Sm8530 Δ*SMb20449, ⊿SMb20447;* Sm^r	This study
IBR519	Sm8530 Δ*SMb20450, ⊿SMb20447;* Sm^r	This study
IBR520	Sm8530 Δ*SMb20452, ⊿SMb20447;* Sm^r	This study
IBR521	Sm8530 Δ*SMb20450, ⊿SMb20451, ⊿SMb20452;* Sm^r	This study
IBR2001	Sm8530 ΔSMb20450, *expA3::Tn5*; Nm^r	This study
E. coli strains		
DH5α	*Φ80lacZ⊿M15,⊿(lacZYA-argF),U169,recA1,endA1, hsdR17(rK–,mK+),phoA,supE44-λ–thi-1,gyrA96, relA1*	Gibco-BRL

Intracellular c-di-GMP levels quantification.

c-di-GMP was extracted and prepared according to the protocol described in (Perez-Mendoza *et al.*, 2014). Briefly, formaldehyde at a final concentration of 0.19% was added to liquid cultures to stop the metabolism of the cells. Cells were harvested by centrifugation, washed with 1 ml cold dH_2O, and resuspended in 0.5 ml of cold dH_2O. Samples were heated for 5 min at 95ºC. 100% iced ethanol was added to a final concentration of 65%. Nucleotides were extracted by vortexing the samples for 30 seconds and centrifuging for 3 min at 13.000 rpm. Supernatants were evaporated in a speed-vacuum system at 50ºC, and nucleotides were resuspended in 300 µl of $AcNH_4$ 10mM (pH 5.5). Samples were filtered with a 0.45 µm GHP membrane (GHP Acrodisc, PALL).

c-di-GMP samples were analysed by reverse phase-coupled HPLC-MS/MS at the mass spectrometry service from Universidad de Jaén. To assure the detection of the nucleotides, the concentration of each sample was raised by adding synthetic c-di-GMP (Axxora) to a final concentration of 250nM. High performance liquid chromatography was performed on an Agilent 1100 coupled to a 3x125 mm column Waters Spherisorb 5 µm ODS2 (C-18). ESI-MS mass spectra were measured on an Esquire 6000 (Bruker Daltonics) and on a TSQ7000 (Finnigan) mass spectrometer. Matrix-assisted laser desorption ionization spectra were measured on a Reflex III spectrometer (Bruker Daltonics). The area of the ion m/z 540 peak was used to estimate the amount of c-di-GMP in each sample. For quantification, a standard curve was established using synthetic c-di-GMP dissolved in $AcNH_4$ (10 mM pH 5.5) at a range of concentrations (20 nM, 200 nM, 2 mM and 20 mM). After subtracting the basal 250 nM spike, c-di-GMP concentrations in each strain culture were standardised with the total protein content, determined by Bradford assay (Bradford, 1976). Three biological replicates of each strain were measured and values were expressed as pmol c-di-GMP /mg protein ± standard error.

ARN techniques.

ARN extraction and RT-PCR.

Sm8530 was grown in 50 ml TY medium to late exponential phase (OD 0.8-0.9). Cells were harvested by centrifugation and total ARN was extracted with Trizol® according to the manufacturer's specifications (Invitrogen). The DNA remains were eliminated with

RQ1 RNase-Free DNase treatment (Promega) for 30 minutes at 37°C and the ARN was purified once again with Trizol®. Final concentration was measured with a Nanodrop® spectrophotometer and its integrity was verified in a 1.4% agarose gel. First-strand cDNA was prepared as indicated by the iScript™ cDNA Synthesis Kit from Bio-Rad. One microgram of total RNA was used per reaction.

For Reverse-Transcription PCR, 1 µl (50 ng) of cDNA was used for each reaction under the following conditions: 95°C for 3 min, 30 cycles of 95°C/30s, 60°C/30s, 72°C/1 min, and a final step of 72°C for 7 min. Reactions with the same amount of ARN or genomic DNA templates instead of cDNA were used as negative and positive controls respectively. Primers pairs used to amplify the intergenic regions are listed in Table 4.

Samples preparation for microarray.

ARN extraction.

Strains were grown in 5 ml TY for 24h, then centrifuged, washed twice with Minimal Medium (MM) and finally suspended in 1 ml MM. These inoculums were used to start new cultures in flasks with 30 ml MM with an initial OD of 0.05 (three replicates per strain). Flasks were incubated at 28ºC and 180 rpm until they reach an OD of 1.2. Next, they were harvested, washed with 0.1% Sarkosyl (dissolved in TE buffer), frozen with liquid nitrogen, and stored at -80ºC until needed. For ARN extraction, cells were first lysed by using a solution of lysozyme (0.4 mg/ml) and then the Qiagen RNeasy RNA purification kit was used following the manufacturer's instructions. Residual DNA was removed with an RNase-free Dnase I Set (Roche Diagnostics ©), and the quality of the RNA was verified in a 1.4% agarose gel. Final concentration was measured with a Nanodrop® spectrophotometer.

cDNA synthesis and labelling.

cDNA was synthesized from 15 µg ARN using the Superscript II Reverse Trasncriptase (Invitrogen™) and following the manufacturer's instructions. The original RNA was degraded by alkaline hydrolysis and neutralization according to the Thermo Scientific supplied protocol. Then the cDNA was purified with an Illustra CyScribe GFX kit (GE Healthcare Life Sciences) and eluted in 60 µl $NaHCO_3$ 1M.

Amersham fluors Cy3 and Cy5 (GE Healthcare Life Sciences) were used for labelling the cDNA control (wild type) and the mutant, respectively. Fluors were prepared according to the manufacturer's instructions, as well as the labelling process (CyScribe Labelling and Purification kits from GE Healthcare.

Microarray hybridization, image acquisition and data analysis.

Sm14koligo microarrays were purchased from Anke Becker (University of Bielefeld, Bielefeld, Germany). Slides pre-treatment and hybridization were performed according to the manual adapted by herself (Becker A., 2003). For image acquisition a GenePix 4100A Scanner (Axon Instruments, Inc., Foster City, CA, USA) was used. Quantifications of mean signal intensities for each spot were determined using the GenePix Pro 5.0 software (Axon Instruments, Inc.). Array IDs and detailed protocol and experiment conditions were deposited in the ArrayLIMS database (Bioinformatics Resource Facility, CeBiTec, Bielefeld University). Normalization and statistics (Holms test) were carried out using the EMMA microarray data analysis software developed also at the CeBiTec centre (Dondrup *et al.*, 2003). Three independent biological replicates were performed for each experiment. Genes were regarded as differentially expressed if they showed $p \leq 0.05$, $A \geq 6$ and $M \geq 1$ or $M \leq$ ratio. A M value above 1 or below -1 indicated at least a double fold change in gene transcription.

Detection of tagged proteins and *in vivo* phosphorylation assay.

O/N inoculums of *E.coli* strains containing the pMLBAD derivatives and the corresponding empty vectors were used to start new 50 ml cultures with an initial OD=0.05. Once they reach an OD=0.4, arabinose was added (0.2% final concentration), and cultures were induced for two hours. For the *S. meliloti* cultures, they were induced with 2% arabinose at an initial OD=0.4 until they reach an OD=0.8 (four hours approximately). In the case of cumate induction, 5 ml of *S. meliloti* TY cultures were induced overnight with 10-50 µM cumate. Cells were harvested by centrifugation (10 min/6000 x g/10ºC). Pellets were resuspended in 1ml dsH_2O, centrifuged again at 13.000 rpm for 5 min, frozen with liquid nitrogen, and stored in aliquots at -20ºC.

Total protein extraction was performed by suspending the aliquots in the BugBuster® Master Mix lysis buffer (10-20 µl for 1 OD_{600}; Novagen) with protease inhibitors (Complete, Mini, EDTA-free, Roche) and incubating at room temperature for 1 hour. Lysates were centrifuged (20 min/20,000 x g/4ºC). Protein extraction from *S. meliloti* cells

harbouring pQF derivatives were performed by centrifuging the cells and resuspending them in 1x Laemly buffer. Cells were boiled for 10 min and centrifuged for another 10 min at 14000 rpm. 0.2 OD_{600} units of the supernatants and 2 OD_{600} units of the pellets were loaded into a 12% SDS-polyacrylamide gels (Gallagher, 2012) non-supplemented or supplemented with 25 μM Phos-tag™ (NARD Chemicals, Hiroshima) and 50 μM $MnCl_2$. Proteins were transferred to a Protran BA85 nitrocellulose membrane (GE Healthcare Life Sciences), and stained with Red Ponceau to verify the transferences. Membranes were subjected to a western blot analysis (Gallagher *et al.*, 2008) using the following antibodies to detect the tagged proteins according to the manufacturer's instructions: HA HRP (Roche Diagnostics), c-Myc HRP (Santa Cruz Biotechnology Inc.) and Strep-Tactin® AP (IBA©).

Phenotypic analysis.

Motility assays.

Swimming motility was tested on TY medium with 0.3% agar (w/v). Plates were inoculated with bacteria from two days solid cultures, using a sterile toothpick and incubated at 28°C. Swimming halos were measured after 4 days. To avoid the evaporation of the media, swimming plates were introduced in a small expandable polystyrene box with moistened papers on the bottom.

Adhesion to microplates wells.

The adhesion of the different strains to polystyrene microplates was tested following the protocol described by Rinaudi and González (2009) with some modifications. Overnight TY cultures were diluted to an OD_{600nm} of 0.8 and 100 μl aliquots were placed in microplate wells (polystyrene VWR® Tissue Culture Plates 96 flat-bottom wells, sterile). The microplates were covered with a sterile breathable rayon film (VWR®) and encased up-side down inside a covered expandable polystyrene box with moistened papers on the bottom, to avoid culture evaporation. By incubating the microplate up-side down, only those cells that have attached to the well walls will remain. Microplates were incubated during 48h at 30ºC and 140 rpm. After removing the liquid content with a micropipette, wells were air dried first, and then gently washed three times with 150 μl of 0.95% NaCl. Wells were allowed to air dry again, and stained with 150 μl of 0.1% violet crystal for 20

min. The dye was removed, wells were washed three times with distilled water, and air dried. Finally, 150 µl of 95% ethanol was added to the wells, and microplates were placed in a shaker at 100 rpm to facilitate the violet crystal dissolution. Absorbance at 570nm was measured with a microplate reader (Biotek® Synergy HT). At least three biological and eight technical replicates were analysed for each strain. Results were expressed as relative adhesion compared to the wild type Sm8530.

Nodulation assays.

Medicago sativa seeds were surface-sterilised during 20 min, with gently agitation, in a falcon tube containing 20 ml of a 50% bleach solution, and then washed ten times with sterilised distilled water (sdH_2O). Seeds were placed on petri dishes (around 25 seeds per plate) with 1% purified agar and allowed to germinate during 24h in darkness and 24ºC. Once germinated, seedlings were humidified with some drops of sdH_2O and the remains of the seed covers were removed with the aid of tweezers. Seedlings were placed in a Petri dish containing 40 ml of solid BNM medium (one seedling per plate). A hole was made in the side of the plates to allow the growth of the aerial part of the plant. On the other hand, inoculums of the desired strains were gown overnight in liquid TY and washed three times with sterilised distilled water. One ml of a 1:50 dilution of these cultures was used to inoculate each plant. Once inoculated, alfalfa plants were incubated for four weeks in a plant chamber (16h light/8h dark photoperiod, 24ºC). The nodules appearance was registered. At least four biological replicates were analysed for each strain.

Calcofluor binding assays.

The followed protocol, with some modifications, is described by Pérez-Mendoza *et al* (2015). Inoculums of the desired strains were grown O/N in liquid TY, then centrifuged, and pellets were washed three times with minimal medium (MM). Cultures were diluted 1:100 in flasks containing 10 ml of MM supplemented with CF (100 µM final concentration) and incubated between 24-28 hours at 30ºC and 180 rpm. Then, cultures were centrifuged for one hour at 4000 x *g*. Supernatants containing unbound CF were removed. Pellets were suspended in 2 ml of distilled water and placed in 24-well plates. Similar growth of all strains was confirmed by measuring the OD_{600nm}. Fluorescence was

measured with a FluoDiaT70 Microplate Reader (Photon Technology International) by exciting the samples at 365nm and recording the emission at 450 nm. Three biological replicates per strain were analysed. The results were expressed as arbitrary units ± SD.

β-galactosidase activity assays.

Strains used in these experiments carried a chromosomal insertion of a *lacZ-aacC1* cassette in the desired genes. The assays were performed following the protocol described by Miller (1972). Cell cultures with an OD_{600nm} around 0.4 were subjected to the colorimetric reaction. The OD was measured at 420nm and 550nm using a Biospectrometer (Eppendorf AG). Transcriptional activity was calculated in Miller units. At least two biological and three technical replicates were used for each assay. Every experiment was repeated three times. Data was expressed as a relative transcription activity to the wild type background, and subjected to the ANOVA test using $P<0.05$ as a threshold for significant differences.

Exopolysaccharide purification and analysis.

EPS II extraction and quantification.

Overnight TY inoculums were centrifuged and washed three times with YMB medium. This inoculums were used to start new cultures in flasks with 200 ml YMB medium (initial OD = 0.05) that were incubated at 180 rpm and 30ºC for six days. Then, cultures were centrifuged for 15 min at 15000 x g and 4ºC. Supernatants were frozen at -80ºC, lyophilized, and suspended in 20ml of NaCl 0.1M, following by the addition of 10 volumes of cold 100% ethanol. Next, samples were precipitated at -20ºC overnight and centrifuged for 20 min at 5000 x g and 4ºC. Pellets were frozen and lyophilized.

For the EPS II quantification, the colorimetric method described by Dubois *et al* (1951) was followed. Three biological and technical replicates were analysed per each strains. Quantification was given as glucose equivalents, interpolating the OD values into a standard curve.

Purification and analysis of the ML β-glucan.

To purify the new exopolysaccharide, we followed the protocol described by Perez-Mendoza *et al.* (2015) with some modifications. Briefly, a starting TY culture of the desire strain was diluted 1:100 in flasks containing 250 ml of YMB and incubated for three or four days at 28ºC and 180 rpm. The flocks were collected using a sieve with a 90-µm cutoff and subjected to four rounds of washing with boiling Milli-Q and cool down. Then, the EPS was centrifuged for 20 min at 3,220 × g, frozen and lyophilized. Infrared spectra were obtained on a FTIR Bruker IFS66v, using potassium bromide (KBr) discs. The transmittance was measure from 550 to 4000 cm^{-1} at resolution of 4 cm^{-1}. The FITR analysis was carried out by the SIdI service from the Universidad Autónoma de Madrid.

Bacteria attachment to alfalfa roots.

The followed protocol was based on the one described by Pérez-Mendoza *et al* (2015) with some modifications. *Medicago sativa* seeds were surface-sterilised and germinated as explained in a previous section.. Once germinated, seedlings were humidified with some drops of sdH_2O and the remains of the seed covers were removed with the aid of tweezers. Seedlings were placed in line in 14 cm diameter Petri dishes (12 seedlings per plate) containing 100 ml of solid BNM medium. Along the surface medium, the plates had also a rectangular piece of sterilised filter paper in order to facilitate the recovering of the roots (Figure 7). The plates were covered with aluminium paper (only the roots area) and incubated for 48h in a plant chamber (16h light/8h dark photoperiod, 24ºC).

Figure 7. Assembly of alfalfa sheedlings in Petri dihes for the root attachment assay.

Roots were inoculated with 30 µl of a sdH_2O solution containing 10^5 cells of each strain (1:1 ratio) and incubated during 1, 3 or 5 days in the same conditions above mentioned. Dilutions of the inoculum were plated on TY with or without the corresponding antibiotic to confirm that both strains were in the same proportion. At the desired time, roots were separated from the shoots, weighed, and introduced in Eppendorf tubes (three roots per tube, four groups in total). They were washed four times with sdH_2O in order to remove the unbound cells. Finally, 1 ml of TY with 2mM EDTA was added. To facilitate the detachment of the cells, roots were subjected to two cycles of 1 min vortexing and 1 min sonication in a water bath. Serial dilutions were plated on TY with or without the corresponding antibiotic and incubated for 3 or 4 days at 28ºC. CFUs of each strain were counted. At least 4 biological and 3 technical replicates were performed for every assay. Data was expressed as the average (in percentage) of mutant CFUs regarding the total ones.

Bioinformatic analysis.

Gene annotations and localizations were studied using the Sm1021 genome available at the web page from the INRA-CNRS institute (http://iant.toulouse.inra.fr/S.meliloti). The presence of conserved domains in the proteins was identified with the Conserved Domain Database from the NCBI institute (Marchler-Bauer *et al.*, 2015) and the InterPro database from the EMBL-EBI institute (Mitchell *et al.*, 2015). Amino acid sequence aligments were performed with the T-Coffee software (Notredame *et al.*, 2000). The cluster analysis of the microarray data based on the GO annotation was carried out by using the free version of the Blast2go software (Conesa *et al.*, 2005). The homology of the *SMb20447-SMb20452* in other bacteria species was analysed by using the KEGG data base (Kanehisa *et al.*, 2015).

Table 4. Primers used in this project.

Primer pairs	PCR product	Sequence [a]
Construction of mutants		
OB67 OB68	*SMb20450* internal fragment	5' GCGCGCGATCCGAAGTCA 5' GCCGGATGCCCAGAGGTTG
OB94 OB95	*SMb20460* internal fragment	5' GCCGCTCTGTATCGGCTCTCACTA 5' AGGCCAGAAACTCCGGATAGGTCA
OB98 OB99	*SMb20447* internal fragment	5' GCTTCTGGAGCGTACGATAAACAA 5' GCAAAAACGTCACGAGATTCAAAG
OB118 OB119	*SMb20447* up-stream fragment	5' TCGTGAAATGACGGAGGAGC 5' **AGATCT**CGATGGAAGATGAA
OB120 OB121	*SMb20447* down-stream fragment	5'**AGATCT**ATGTTACCGCCGAAGGAGTG 5'GCGCCAGGTCATTTTCGTTT
OB122 OB123	*SMb20448* up-stream fragment	5'ACGAGTTCGCGGTCATTCAGTGG 5'**GCGGCCGC**CTCCCTTACGCAAAAAC
OB116 OB70	*SMb20448* down-stream fragment	5'GAGTCGCGAGGCTGTTCTTCTG 5'GATGGAGGGCGGCAAAAATGTGGT
OB120 OB117	*SMb20449* up-stream fragment	5'**AGATCT**CGATGGAAGATGAA 5'**GGATCC**ACGACGCTCGAGGTCAC
OB91 OB86	*SMb20449* down-stream fragment	5'**GGATCC**GGTCGCGAGGCTGTTTAG 5'**GCATGC**GGAAGCGGAAGGTGGTTC
OB85 OB86	*SMb20450* up-stream fragment	5'**GGATCC**GCGGGCGGGTCTATA 5'**GCATGC**GGAAGCGGAAGGTGGTTC
OB87 OB88	*SMb20450* down-stream fragment	5'**GCATGC**GGCGATCTACAAGAACC 5'**AAGCTT**AAGCAGGTGGCGGAACTT
OB93 OB126	*SMb20451* up-stream fragment	5'CAAGCGCATCGAGGCGAAGGACT 5'**CCCGGG**TCCCGACGGCAAGATC
OB124 OB125	*SMb20451* down-stream fragment	5'**CCCGGG**ACCTCTCGAATTGACG 5'ACTCGAACTATAAGTGAACC
OB87 OB127	*SMb20452* up-stream fragment	5'**GCATGC**GGCGATCTACAAGAACC 5'**AAGCTT**GATTTCCATGTTTTCG
OB128 OB125	*SMb20452* down-stream fragment	5'**AAGCTT**CACCCTGGCGTCGATC 5'ACTCGAACTATAAGTGAACC
RT-PCR		
OB98 OB74	783 bp fragment between *SMb20447* and *SMb20448*	5' GCTTCTGGAGCGTACGATAAACAA 5' **GGTACC**GGGGGACACGGGAAAGT
OB116 OB117	404 bp fragment between *SMb20448* and *SMb20449*	5' GAGTCGCGAGGCTGTTCTTCTG 5' **GGATCC**ACGACGCTCGAGGTCAC
OB85 OB86	885 bp fragment between *SMb20449* and *SMb20450*	5'**GGATCC**GCGGGCGGGTCTATA 5'**GCATGC**GGAAGCGGAAGGTGGTTC
OB87 OB88	806 bp fragment between *SMb20450* and *SMb20451*	5'GCATGCGGCGATCTACAAGAACC 5'AAGCTTAAGCAGGTGGCGGAACTT

Primer pairs	PCR product	Sequence [a]
OB111 OB110	414 bp fragment between *SMb20451* and *SMb20452*	5' CCTGCTTTCTGAAATAACGATAGG 5' GAACTGCAGCCCCGTGAGAT
Complementation/Overexpression		
OB73 OB74	2392 bp fragment containing *SMb20447*	5' **GGATCC**CACGCGACCTATGGAC 5'**GGTACC**GGGGGACACGGGAAAGT
OB120 OB117	1044 bp fragment containing*SMb20448*	5'**AGATCT**ATGTTACCGCCGAAGGAGTG 5'**GGATCC**ACGACGCTCGAGGTCAC
OB116 OB86	1768 bp fragment containing *SMb20449*	5'GAGTCGCGAGGCTGTTCTTCTG 5'**GCATGC**GGAAGCGGAAGGTGGTTC
OB91 OB92	2777 bp fragment containing *SMb20450*	5'**GGATCC**GGTCGCGAGGCTGTTTG 5' **GGATCC**GGGCCGTCGAAGAGAC
OB87 OB110	1207 bp fragment containing *SMb20451*	5' **GCATGC**GGCGATCTACAAGAACC 5' GAACTGCAGCCCCGTGAGAT
OB124 OB84	764 bp fragment containing *SMb20452*	5' **CCCGGG**ACCTCTCGAATTGACG 5' CGCCGGACGAGGCGCGACTTTA
OB100 OB101	2129 bp fragment containing *SMb20391*	5'**GGATCC**GGCGCAGGTTCGTTTCTC 5' CGCTCCCGCACGCTCTAT
Point mutations		
OB129 OB130	SMb20447^{E215A}	5'GTCTAGCCGGGGACGCGTTCGCGGTC ATTCAGT 5'CTGAATGACCGCGAACGCGTCCCCGG CTAGACG
OB149 OB150	SMb20452^{S61A}	5' GTTCCTCAATGCCTCCGGAATCAAC 5' TGCAGCCCCGTGAGATCC
Protein tags		
OB131 OB132	*SMb20447* with 5' StrepII tag	5'**GGTACC**ATGTGGAGCCACCCGCAGTT CGAAAAA ATGCCGGACGGCGATACG 5'**AAGCTT**CATGTTCGCTGAAGA
OB133 OB127	*SMb20451* with 5' HA tag	5'**GGTACC**ATGTACCCATACGACGTACCA GATTACGCTATGACAACCTTATATGG 5'**AAGCTT**GATTTCCATGTTTTCG
OB134 OB135	*SMb20452* with 5' c-Myc tag	5'**GGTACC**ATGGAGCAGAAACTCATCTC AGAAGAGGATCTGATGGAAATCAAGG AAGAC 5'**AAGCTT**TCAGCTTAGCCTCAGAT
OB144 OB145	*SMb20447* NdeI-NsiI fragment	5' **CATATG**CCGGACGGCGATACG 5' **ATGCAT**TGTTCGCTGAAGACCTGCC
OB138 OB142	*SMb20451* NdeI-NsiI fragment	5' **CATATG**ACAACCTTATATGGACTG 5'**ATGCAT**ATTCGAGAGGTCGGGGATGG
OB140 OB143	*SMb20452* NdeI-NsiI fragment	5' CATATGGAAATCAAGGAAGACGAC 5'ATGCATGCTTAGCCTCAGATCGACG

RESULTS

Phenotypic screening of mutants affected in genes coding for putative diguanylate cyclases and/or phosphodiesterases in *Sinorhizobium meliloti.*

Sinorhizobium meliloti possesses 20 ORFs annotated as putative diguanylate cyclases (DGCs) or phosphodiesterases (PDEs). They are distributed among the three replicons, and some present additional sensory or response domains like GAF, PAS and REC (Table 5). Thanks to the extensive mutant library elaborated by Anke Becker and collaborators (Pobigaylo *et al.*, 2006), we could make use of a *S. meliloti* 2011 mini-Tn*5* set of mutants affected in eleven different genes with the aim of performing a phenotypic screening (Supplementary Table 1). However, it is known that the strain Sm2011, as its very close Sm1021, presents an insertion sequence in *expR*, a global transcription regulator that forms part of the Sin/ExpR quorum sensing. This system regulates important physiological processes like motility, exopolysaccharide (EPS) production, or nitrogen fixation

Table 5. ORFs in Sm1021 annotated as putative DGCs and/or PDEs. Additional domains are also indicated. Mutants used for the phenotypic screening are listed in the second column.

Gene	Mutant ID	GGDEF	EAL	Additional Domains
SMa0137	A1	1	1	CHASE
SMa1548	A2	1	1	5 PAS
SMa2301		1		
SMa0369			1	
SMb20389	B1	1		GAF
SMb20447	B2	1	1	
SMb20523		1		
SMb20900	B4	1	1	PAS
SMb21517			1	
SMc00887	C1	1	1	
SMc00992		1	1	
SMc00074		1	1	PAS
SMc00033	C4	1	1	
SMc00038	C5	1	1	GAF / PAS
SMc01464		1		
SMc04015		1		
SMc03942	C8	1	1	PAS
SMc03178	C9	1	1	PAS / CHASE
SMc03141	C10		1	
SMc01370 (PleD)		1		2 REC

(Pellock *et al.*, 2002, Hoang *et al.*, 2004). Since DGCs/PDEs activities are often associated with quorum sensing, motility, or EPS phenotypes (Srivastava & Waters, 2012, Hengge, 2009), we considered it would be more appropriate to work with a strain that had a complete and functional Sin/ExpR system. Thus, we transduced the mTn*5* mutations into strain *S.meliloti* 8530 (Sm8530).

The phenotypic screening consisted in three different assays: swimming motility, adhesion to polystyrene micro plates, and nodulation fitness in alfalfa. No significant differences were found in the nodulation assays (data not shown). Regarding the adhesion capacity, the great variability between experiments did not allow us to conclude if any of the mutants behaved different than the wild type. In contrast, we could detect several mutants with higher motility than the wild type on the swimming assays. Among them, the mutant affected in gene *SMb20447* showed also a slight reduction in mucoid colony morphology compared to the parental strain. However, not only we could not complement these phenotypes, but the mutant carrying a copy of *SMb20447 in trans* (pJL2108) presented an even drier phenotype. To further characterize the role of the protein in the dryer and motile phenotypes, we constructed a strain with an unmarked in frame deletion in the gene (Δ*SMb20447*). Surprisingly, this strain showed a wild type phenotype regarding colony mucoidy and motility. A closer look into the *SMb20447* genomic region, located in the pSymB megaplasmid, revealed the presence of five downstream genes, from *SMb20448* to *SMb20452*, displaying the same orientation (Figure 8), and suggesting that the observed phenotypes in the mTn*5* mutant could be due to a polar effect of the mini-transposon over the downstream genes.

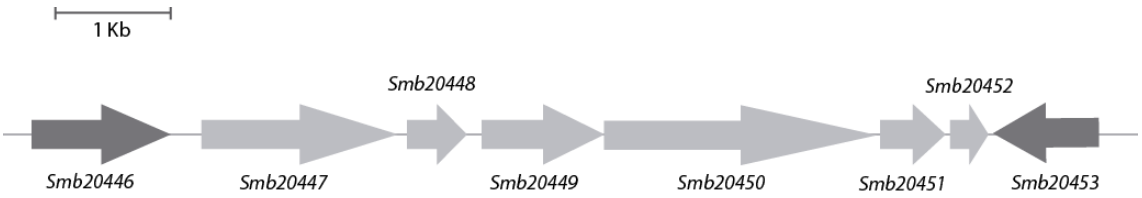


Figure 8. Genetic localization of *SMb20447* in the pSymB megaplasmid.

SMb20447 is active as a DGC.

As mentioned before, a deletion in gene *SMb20447* did not produce any remarkable phenotype. However, its overexpression (pJL2108) displayed a strong brightness on TY plates supplemented with calcofluor (Figure 11). SMb20447 is annotated as a cytosolic protein with both GGDEF and EAL domains. A closer look to the amino acid sequence revealed that the first glycine of the GGDEF domain was in fact an alanine (AGDEF). Nevertheless, there are DGCs with this motif version that have been proved to be active (Hunter *et al.*, 2014). An alignment with other well-known DGCs and PDEs from different bacteria showed that SMb20447 presents all the critical amino acids required for either synthesis or degradation of ci-di-GMP (Figure 9) (Romling *et al.*, 2013, Bobrov *et al.*, 2011).

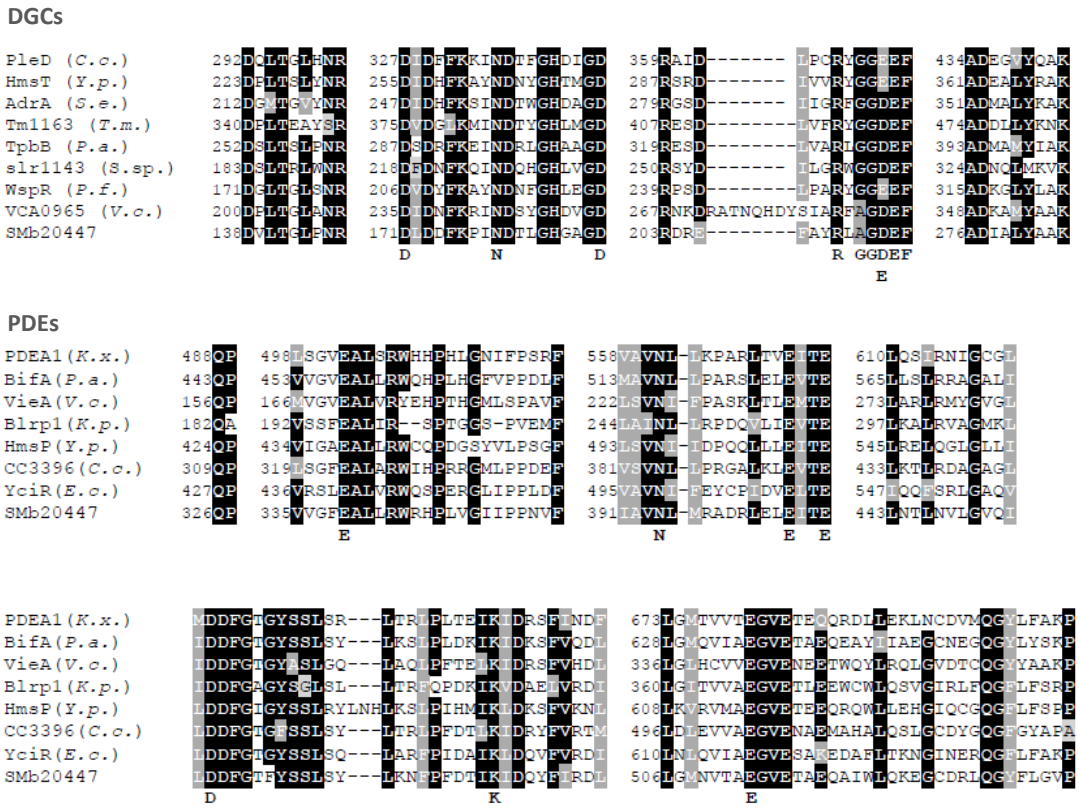


Figure 9. Amino acid sequences alignment of SMb20447 with the GGDEF and EAL domains of known functional DGCs and PDEs. White letters over black or grey backgrounds indicate identical or conservative changes, respectively. Black letters over grey backgrounds denote semi-conservative changes. Bold letters at the bottom of the alignment are those residues described as critical for the domain activity. Bacteria species are indicated between brackets: *Caulobacter crescentus* (C.c.), *Yersinia pestis* (Y.p.), *Salmonella enterica* (S.e.), *Thermotoga maritima* (T.m.), *Pseudomonas aeruginosa* (P.a.), *Synechocystis sp.* (S.sp.), *Pseudomonas fluorescens* F113 (P.f.), *Vibrio cholerae* (V.c.), *Komagataeibacter xylinus* (K.x.), *Klebsiella Pneumoniae* (K.p.), *Escherichia coli* (E.c.).

Additional evidences associated with active DGCs were found. First, the overexpression of *SMb20447* in Sm8530 produced a 37% decrease in swimming motility (Figure 10A). Secondly, intracellular levels of c-di-GMP in Sm8530 overexpressing *SMb20447* were remarkably high compared to the wild type carrying the empty vector (Figure 10B).

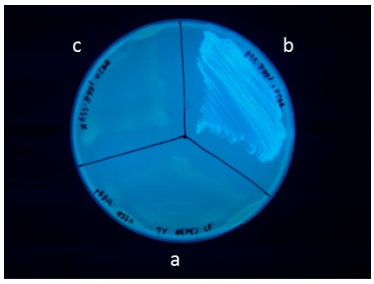


Figure 11. Strains streaked on TY plates supplemented with CF and observed under UV light. **(a)** Sm8530 pBBRMCS-5. **(b)** Sm8530 overexpressing *SMb20447* (pJL2108). **(c)** Sm8530 overexpressing *SMb20447*^{E215A} (pJL2108E215A)

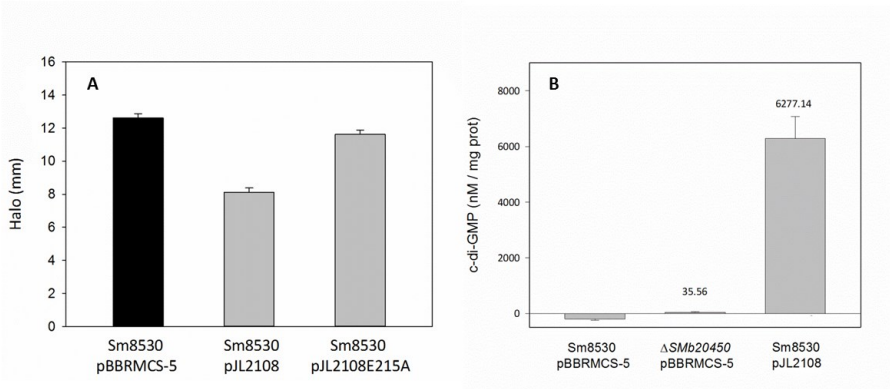


Figure 10. (A) Swimming motility of Sm8530 overexpressing *SMb20447* (pJL2108) or its AGDAF version (pJL2108E215A) compared to the same strain carrying the empty vector. **(B)** Intracellular c-di-GMP levels in Sm8530 and Δ*SMb20450* bearing an empty vector and Sm8530 overexpressing *SMb20447* (pJL2108). The negative value of the control strain might be due to technique limitations when it comes to measure basal levels.

In order to confirm that the observed phenotypes were due to the activity of the AGDEF domain, we carried out a point mutation, replacing the critical glutamic acid by and alanine (SMb20447^{E215A}). The overexpression of this new AGDAF version (pJL2108E215A) in Sm8530 did not produce the calcofluor (CF^{+}) phenotype (Figure 11) and almost restored the wild type swimming motility (Figure 10A). In conclusion, SMb20447 is active as a DGC when is overexpressed, despite it possesses an atypical

AGDEF domain and a well conserved EAL motif. The increment of c-di-GMP as a result of this activity decreases the swimming motility and produces a strong CF^+ phenotype.

A *SMb20450* mutant is less mucoid and displays a blue brightness under UV light.

Considering that the transposon insertion in *SMb20447* could be affecting the downstream genes, we decided to put the target on *SMb20450*, which is located 1767 bp downstream of *SMb20447* and codes for a large putative regulatory protein (786 amino acids) with specific phosphatase and HAMP domains. First, we made an insertion mutation in this gene (*SMb20450::Nm*). This strain displayed a drier phenotype and a blue brightness under UV light in plates supplemented with CF, very similar to what we observed with the overexpression of *SMb20447*. An introduction of a complete copy of *SMb20450 in trans* (pJL2111) restored the wild type phenotype (data not shown). We decided then to make a new mutant affected in this gene, but this time by an in-frame deletion. By this way we would have a clean and irreversible mutation with no antibiotic marker, which would be more convenient in case we wanted to make secondary mutations in other genes. Δ*SMb20450* strain presented the same phenotype as the *SMb20450::Nm* (Figure 12), and it was also complemented by introducing a wild type copy of *SMb20450* (Figure 16C).

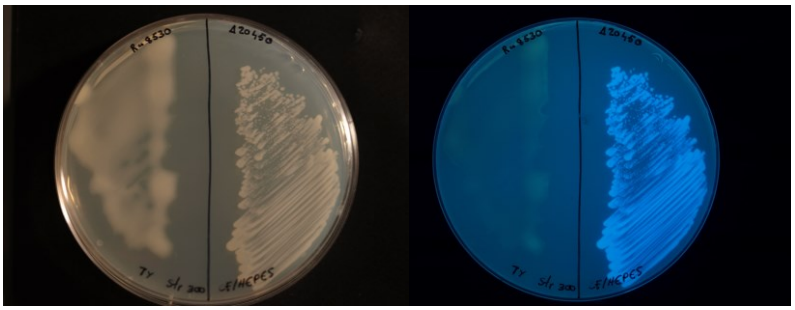


Figure 12. Wild type strain Sm8530 (left side of the plate) and Δ*SMb20450* mutant (right side) streaked on a TY plate supplemented with CF, and illuminated under bright light (left plate) and UV light (right plate).

In addition to the less mucoid phenotype, some other evidences indicated that Δ*SMb20450* strain was defective in EPS II production. First, liquid cultures did not aggregate in static conditions (Figure 13B). Second, they lost almost completely their adhesion capacity to polystyrene microplates, similarly to the defective EPS II strain

Sm1021 (Figure 13A). And last, quantification of total EPS II in the supernatants revealed that the mutant produced 50% less glucose equivalents compared to the wild type Sm8530.

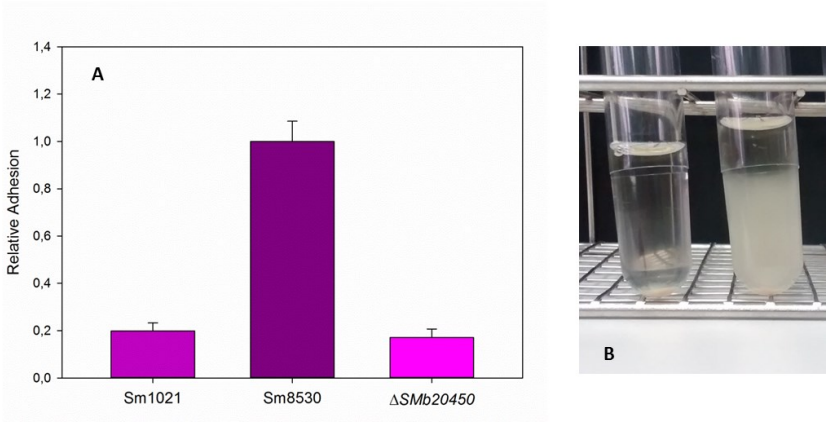


Figure 13.(A) Relative adhesion to polystyrene microtiter plates of strain Δ*SMb20450* compared to the wild type Sm8530 and to the EPS II defective strain Sm1021. **(B)** Static liquid cultures of Sm8530 (left tube) and strain Δ*SMb20450* (right tube).

The Δ*SMb20450* strain produces a cellulose-like exopolysaccharide that depends on the presence of the global regulator ExpR.

Of the two EPS originally described in *S. meliloti*, only succinoglycan (EPS I) binds to calcofluor (CF), giving a green-yellowish brightness under UV light. Instead, the Δ*SMb20450* strain displayed a deep blue brightness that was retained even if key genes necessary for the production of EPS I or both EPS I and EPS II were mutated (Figure 14A). This suggested that another exopolysaccharide was present. In order to roughly characterize the nature of this new polymer, we used different dyes that bind to other well-known polysaccharides: Congo red (CR) interacts with basic or neutral β-D-glucans like cellulose (Wood, 1980); aniline blue, on the other hand, is more specific for glucans mainly formed by β-1,3-glucosidic linkages, like curdlan (Nakanishi *et al.*, 1976). As shown in Figures 14B and 14C, the derivative of Δ*SMb20450* strain defective in EPS I and EPS II was stained with CR but not with aniline blue.

It is worthy to distinguish that a *SMb20450* mutation in the strain Sm1021 does not produce the phenotype above mentioned (Figure 14D). This strain is genetically identical to Sm8530 except for the presence of an insertion sequence that inactivates *expR*, a global transcriptional regulator that controls the transcription of the *exp* genes (EPS II) among other important physiological processes, and also forms part of the Sin/ExpR quorum sensing system (Pellock *et al.*, 2002).

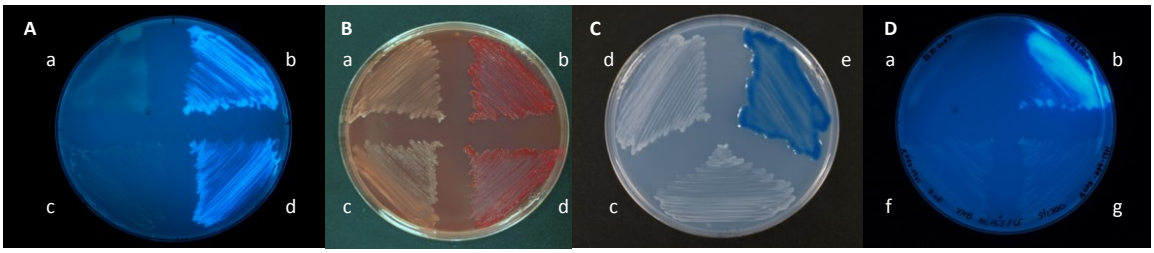


Figure 14. Different EPS defective versions of Sm8530, Sm1021 and the Δ*SMb20450* strains streaked on TY or YMB plates supplemented with CF (**A** and **D**), CR (**B**), or aniline blue (**C**). Lower case letters indicates the following strains: **(a)** Sm8530 EPS I⁻, **(b)** Δ*SMb20450* EPS I⁻, **(c)** Sm8530 EPS I⁻/EPS II⁻, **(d)** Δ*SMb20450* EPS I⁻/EPS II⁻, **(e)** Curdlan producer strain C58 from *Agrobacterium tumefaciens*, **(f)** Sm1021 EPS I⁻, **(g)** Sm1021 *SMb20450::Nm* EPS I⁻.

In liquid medium, especially YMB, the Δ*SMb20450* strain formed white fibrils that disappeared when the culture was treated with commercial cellulase (Figures 15A and 15B). The fibrils could be CF stained and presented a cottony aspect when observed at the microscope under UV light (Figures 15C and 15D). Bacteria seemed to be enclosed in these structures, and were released in significant amounts when the fibrils were smashed.

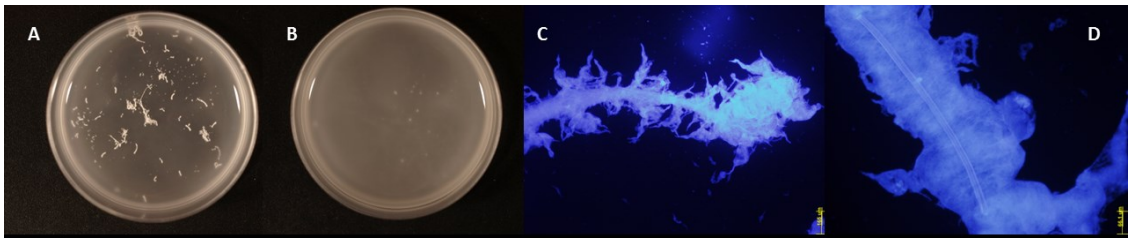


Figure 15. Fibrils formed by Δ*SMb20450* strain **(A)** were degraded when treated with commercial cellulase **(B)**. Fibrils could be stained with CF and observed at the microscope under UV light **(C** and **D).**

SMb20391 is the synthase responsible for the production of the cellulose-like polymer, recently characterized as a ML β-glucan.

Given the similarity of this new polymer to cellulose, we searched for genes in the Sm8530 genome that were annotated as cellulose synthases. Two candidates were found: *SMb20460* and *SMb20391*. Both are located on the pSymB megaplasmid, 14 kb downstream and 52 kb upstream gene *SMb20450*, respectively. To find out if they were implicated in the synthesis of the polymer, insertion mutations were made in each gene under the Δ*SMb20450* background. No changes were observed when *SMb20460* was disrupted, except for the smoother aspect of the fibrils (Figure 16A). In the case of *SMb20391*, however, the blue brightness and fibril production was abolished (Figure 16C). The introduction of an intact copy of *SMb20391* (pJL2114) in the double mutant *SMb20450/SMb20391* restored the phenotype of the Δ*SMb20450* strain. Thus, SMb20391

seems to be the synthase required for the production of the exopolysaccharide. A closer look into the amino acid sequence of this protein revealed the presence of characteristic domains both from cellulose and curdlan synthases (Stasinopoulos *et al.*, 1999, Romling, 2002).

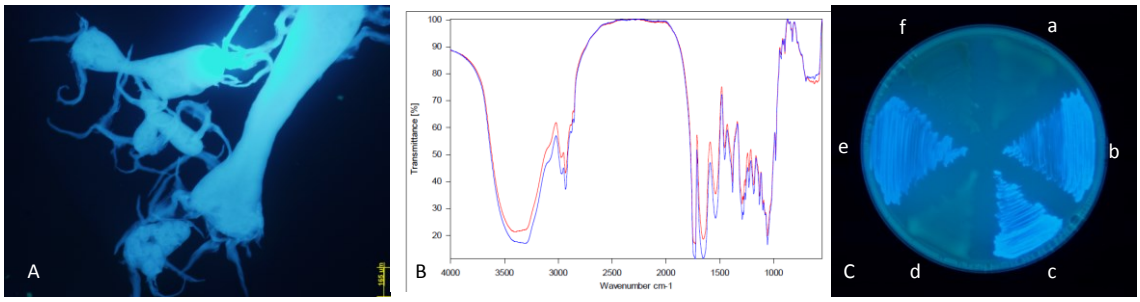


Figure 16. A. CF stained fibrils from double mutant *SMb20450/SMb20460*. **B.** FTIR transmission spectrums from the ML β-glucan (blue line) and the polymer produced by strain Δ*SMb20450* (red line). **C.** TY plate supplemented with CF showing **(a)** Sm8530, **(b)** Δ*SMb20450*, **(c)** Sm8530 pJBpleD*, **(d)** Δ*SMb20450* pJL2111, **(e)** Δ*SMb20450* EPS I$^-$/EPS II$^-$, **(f)** *SMb20450/SMb20391* double mutant**.**

Concurrently, Pérez-Mendoza *et al.* (2015) described a new exopolysaccharide produced by Sm8530 in response to the overexpression of the DGC PleD from *Caulobacter crescentus* (strain Sm8530 pJBpleD*). This EPS was also CR positive, displayed a blue brightness on CF plates (Figure 16C), and formed flocs in liquid media. They characterized it as a linear β-glucan with alternate (1→3), (1→4) linkages (ML β-glucan), and designed SMb20391 (later BgsA) as the corresponding synthase. FTIR spectrums of this ML β-glucan and the polymer produced by strain Δ*SMb20450* proved to be practically identical, pointing out they were the same EPS (Figure 16B). These results indicated that SMb20450 negatively regulates BgsA, the synthase responsible for the production of the ML β-glucan.

SMb20447 is negatively regulated by SMb20450, and acts as a c-di-GMP source for the ML β-glucan production.

Considering that both the overexpression of *SMb20447* and the mutation in *SMb20450* produced the same CF$^+$ phenotype, we wondered whether the activity of these proteins was related to each other. A strain carrying deletions in both genes showed no CF fluorescence (Figure 17) and recovered the mucoid aspect of Sm8530. The introduction of a *SMb20447* copy in *trans* (pJL2108) in this strain restored the Δ*SMb20450* strain

phenotype regarding CF (Figure 17) and CR (data not shown). Thus, it seems that SMb20450 and SMb20447 belong to the same pathway, being the former protein a negative regulator over the later. However, we did not know whether SMb20447 was specifically required for the production of the ML β-glucan or it could be replaced by another DGC that provided a c-di-GMP source. To approach this question, we introduced a plasmid bearing a constitutively active version of PleD (pJBpleD*) in a Δ*SMb20447* strain and the double mutant *SMb20450/ SMb20447*. In both cases CF (Figure 17) and CR tests were positive, suggesting that the role of SMb20447 is just to provide c-di-GMP to BgsA, and thus it can be replaced by another DGC.

On the other hand, intracellular ci-di-GMP levels in the Δ*SMb20450* strain were slightly higher than in the wild type strain, confirming the negatively regulation of SMb20450 over SMb20447 (Figure 10B). This increment is minimal, however, compared to the levels found in Sm8530 overexpressing *SMb20447*. We considered that the intracellular ci-di-GMP concentration in the Δ*SMb20450* strain would correspond to physiological levels, whereas the overexpression of *SMb20447* is totally out-regulated due to the high copy number of the vector.

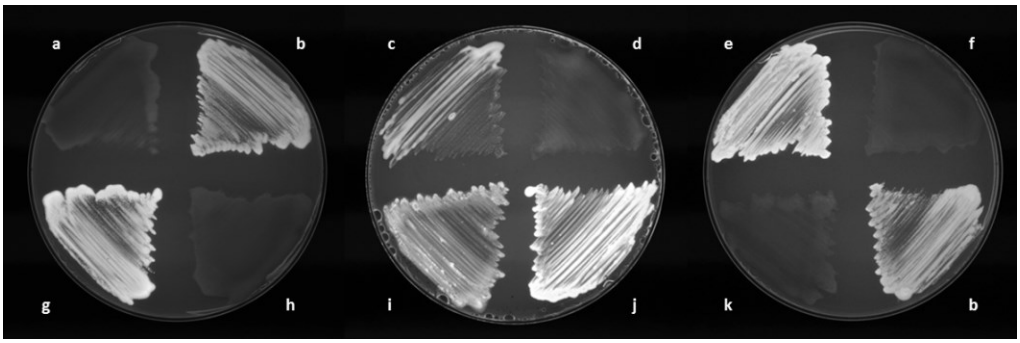


Figure 17. Different strains streaked in TY plates supplemented with CF. **(a)** Δ*SMb20450* pJL2108E215A **(b)** Sm8530 pJL2108 **(c)**. Δ*SMb20450* **(d)** *SMb20450/SMb20447* double mutant **(e)** *SMb20447::Gn* pJBpleD* **(f)** Sm8530 **(g)** Δ*SMb20450* pJL2108 **(h)** Sm8530 pJL2108E215A **(i)** *SMb20450/SMb20447* pJBpleD* **(j)** *SMb20450/SMb20447* pJL2108 **(k)** *SMb20447::Gn.*

Surprisingly, the overexpression of the AGDAF version of SMb20447 (pJL2108E215A) in a Δ*SMb20450* strain restored the wild type phenotype regarding CF and mucoid phenotypes, whereas it did not produce any remarkable change under a wild type background (Figure 17). This last result made us consider the hypothetical dual role of SMb20447 as a DGC and PDE. Actually, we are looking forward to receiving new c-di-

GMP measurements from strains Sm8530 and Δ*SMb20450* overexpressing the AGDAF version of *SMb20447* that hopefully will give us new insights.

The region *SMb20447-SMb20452* forms a single transcription unit.

The close disposition of genes *SMb20447* and *SMb20450* along with the putative polar effect observed in the *SMb20447::Tn5* mutant made us consider they could be transcribed in a single operon. There are six genes in this region, from *SMb20447* to *SMb20452*, located very close to each other (Figure 18A).

To confirm this idea, total ARN from Sm8530 was extracted, transformed to c-DNA, and subjected to RT-PCR. We selected five primer pairs covering the corresponding intergenic regions (Table 4, and Figure 18A). As shown in Figure 18B, there was amplification in all cases, confirming that the six genes are co-transcribed. To assure the amplification was only due to the presence of c-DNA, additional positive and negative controls were included with genomic DNA or RNA as templates, respectively. Additionally, we also performed an RT-PCR of the regions between genes *SMb20446-SMb20447* and *SMb20452-SMb20453* without obtaining any amplification (data not shown). This result suggests that the operon is only comprised by the six genes above mentioned.

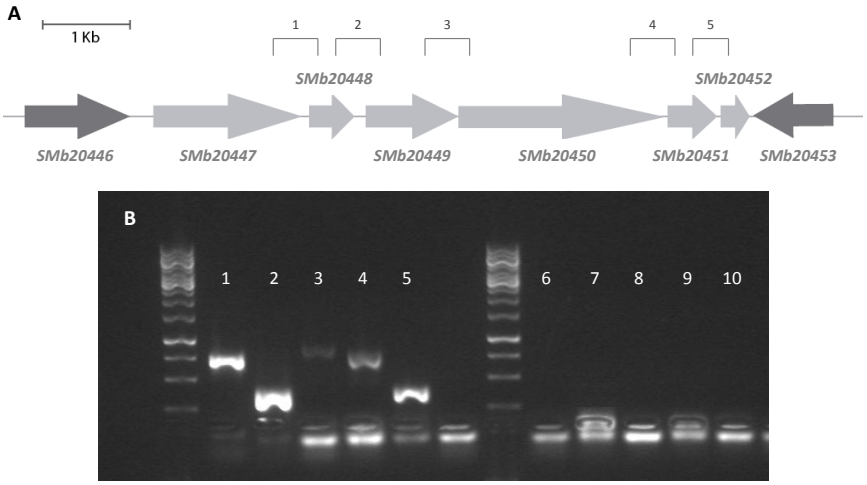


Figure 18. Genetic map of the *SMb20447-SMb20452* region **(A)** and RT-PCR of the corresponding intergenic fragments **(B, lines 1-5).** The same reactions using ARN as a template were also carried out as negative controls **(B, lines 6-10)**.

Transcription of *SMb20391* is regulated by ExpR and enhanced by c-di-GMP.

As outlined before, a *SMb20450::Nm* mutation in the strain Sm1021 does not produce the blue phenotype on CF plates, suggesting that the ML β-glucan synthesis is dependent on the global transcription regulator ExpR. To confirm this evidence we made *lacZ* fusions to *SMb20391* (the ML β-glucan synthase) and measured the transcription activity in strains Sm8530 ($ExpR^+$) and Sm1021 ($ExpR^-$). As expected, the transcription in the $ExpR^+$ background is around 10 fold higher compared to the $ExpR^-$ strain (Figure 19A). This result confirmed what had been previously found by Hoang *et al.* (2004) in microarray analysis and most recently by Perez-Mendoza *et al.* (2015).

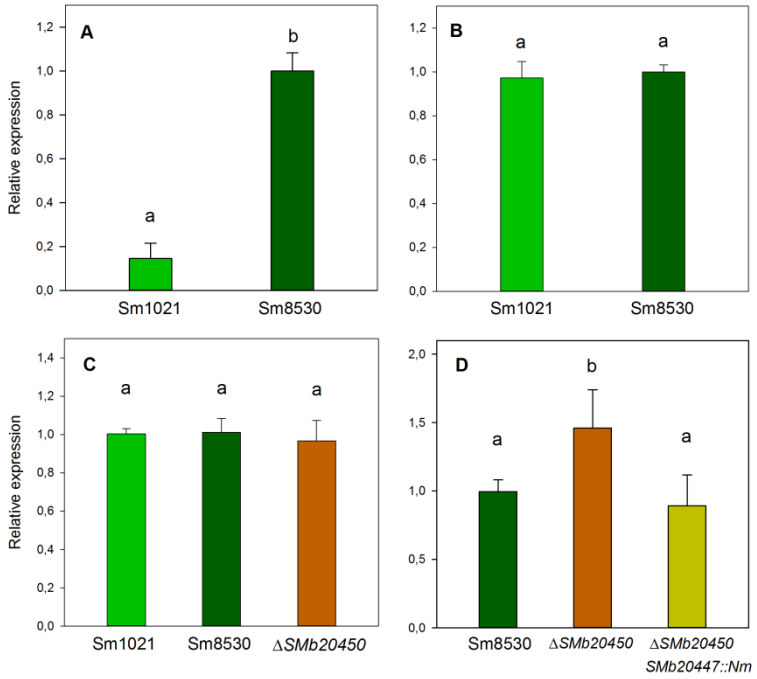


Figure 19. Relative expression of genes *SMb20391* **(A** and **D)**, *SMb20450* **(B)** and *SMb20447* **(C)** fused to a promoterless *lacZ* gene in different genetic backgrounds. Transcription levels in Sm8530 were taken as a reference. Errors bars represent standar deviations; different lower case letters above the bars denote statistical significant differences (analysis of variance with $p<0.05$).

On the other hand, we have seen that SMb20447 is active as a DGC and is needed for the production of the ML β-glucan, at least as a c-di-GMP source. However, we did not know whether c-di-GMP was also regulating the transcription of *SMb20391* or either was acting at other levels. For this purpose, we measured the transcription activity of *SMb20391-lacZ* fusions in different genetic backgrounds (Figure 19D). In a Δ*SMb20450*

background the transcription activity of *SMb20391* displays a 50% increment, but it drops again to wild type levels when *SMb20447* is not present. Thus, c-di-GMP seems to enhance the transcription of *SMb2391* although ExpR is clearly the main regulator at this level.

Since the expression of the ML β-glucan synthase is clearly controlled by ExpR, one might suspect that also the operon *SMb20447-SMb20452* is submitted to this global regulator. Figures 19B and 19C show the relative expression of genes *SMb20447* and *SMb20450* in ExpR$^+$ and ExpR$^-$ backgrounds. No statistical differences were found, which indicates that the operon is expressed independently from ExpR. Additionally, the expression of *SMb20447* does not increase in the Δ*SMb20450* background (Figure 19C), suggesting that the negative regulation of SMb20450 concerning SMb20447 might be at the post-transcriptional level.

The six genes from the operon *SMb20447-SMb20452* participate in the ML β-glucan production as positive or negative regulators.

Once we knew that the *SMb20447-SMb20452* region constitutes and operon, we wondered whether they were all implicated in the ML β-glucan regulation. For this purpose, we made in-frame deletions in the remaining genes and analysed their phenotypes in TY plates supplemented with CF. Δ*SMb20449*, Δ*SMb20451* and Δ*SMb20452* strains displayed a blue brightness and less mucoid phenotype, although with different intensities. Wild type phenotype was restored when a copy of the corresponding gene was introduced *in trans* (plasmids pJL2110, pJL2112 and pJL2113, Table 2), except for the Δ*SMb20451* strain, where only a copy of *SMb20452* could revert the wild type phenotype (Supplementary Table 2). Although we were careful of making in-frame deletions to avoid polar effects, any error during the double homologous recombination could have affected the expression of *SMb20452* in the Δ*SMb20451* strain. However, the strain Δ*SMb20451* presented an intermediate CF$^+$ and mucoid phenotype between the wild type and the Δ*SMb20452* strain, so we could not discard either the existence of an internal promoter within *SMb20451* sequence that disappears with the deletion and thus alters the transcription of the downstream gene *SMb20452*. Supporting this theory, a global mapping of transcription start sites in Sm1021 predicts the presence of an internal promoter in *SMb20451* (Schluter *et al.*, 2013).

On the other hand, we also analysed the overexpression of each gene in a wild type background (plasmids from pJL2108 to pJL2113, Table 2). *SMb20447*, *SMb20448* and *SMb20451* genes produced a CF positive and less mucoid phenotypes when overexpressed, again with different intensities (Supplementary Table 2).

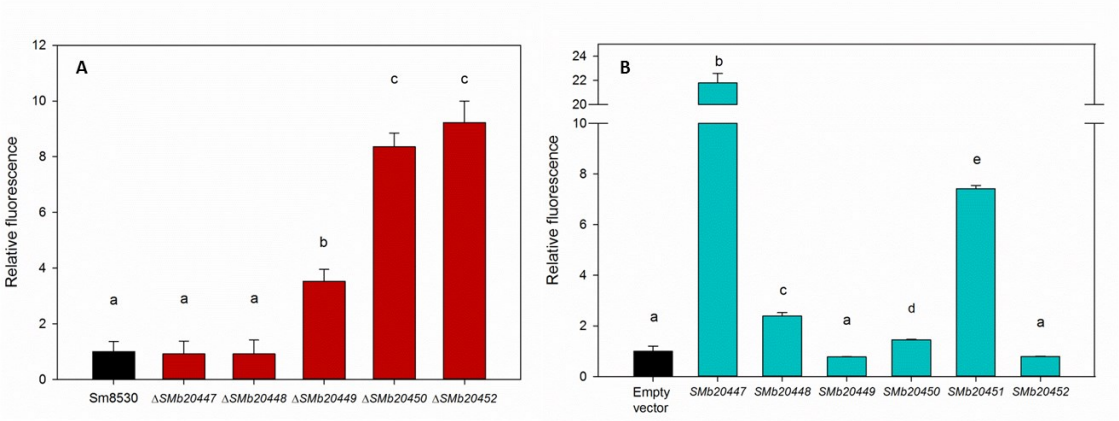


Figure 20. (A) CF binding assay showing the fluorescence of the different deletion mutants from the operon compared to the wild type Sm8530. **(B)** Fluorescence produced by the overexpression of each gene of the operon in a wild type background compared to the wild type strain carrying the empty vector. Different letters above the bars indicate statistical significant differences ($p<0.05$).

With the aim of quantify the contribution of each gene in the ML β-glucan production, we measured the fluorescence of all the deletions mutants and the overexpressions in liquid Minimal Medium (MM) supplemented with CF. Strain Δ*SMb20451* was not included in this assay due to its probable polar effect over *SMb20452* above mentioned. As Figure 20A shows, strains Δ*SMb20450* and Δ*SMb20452* produced around nine fold more fluorescence than the wild type, whereas strain Δ*SMb20449* was also more fluorescent but to a lesser extent than the formers. This was consistent with the phenotypes observed on CF plates. On the other hand, Figure 20B shows that the overexpression of *SMb20447* gave the highest fluorescence (twenty two times more than the wild type), followed by the overexpression of *SMb20451* and *SMb20448* (eight and more than two fold compared to the wild type, respectively). In this case, the results were also consistent with the phenotypes observed on CF plates, with one exception. The fluorescence given by the overexpression of *SMb20450* was significantly different according to the statistical analysis. However, it did not produce any CF^+ phenotype on plates and neither any fibril could be detected on liquid medium. Considering that the difference detected in the fluorescence assay was minimal (less than two fold), we assumed that the significant difference was due to the very low variability among the replicates.

These results indicate that the six genes of the operon are implicated in the regulation of the ML β-glucan to a greater or lesser extent: SMb20449, SMb20450 and SMb20452 act as negative regulators, whereas SMb20447, SMb20448 and SMb20451 are positive regulators. It is worthy to highlight that those genes producing a CF^+ phenotype when overexpressed, do not have any effect when they are mutated, and vice versa. Thus, we speculated that a strict *SMb20451* mutant should not have any effect, since its overexpression displayed a strong CF^+ phenotype.

The operon *SMb20447-SMb20452* and ExpR are the main actors involved in the regulation of the ML β-glucan.

The role of the operon *SMb20447-SMb20452* and the global regulator ExpR in the production of this new exopolysaccharide was clear, but we wondered whether other genes could be also implicated. In collaboration with Professor Juan Sanjuán's group (Estación Experimental del Zaidín, CSIC, Granada), we performed a general mutagenesis in Sm8530 by using a Tn*5* transposon. The objective was to select colonies with a Congo red positive phenotype (CR^+), which is indicative of the ML β-glucan production. Transposants were plated on minimal medium supplemented with CR. In the end, ten colonies remained as candidates. Figure 21 shows a Southern Blot of these colonies and a wild type control. Genomic DNA was digested with StuI, a restriction enzyme that generates a 12kb fragment containing the complete operon of interest. The probe used was an 806 bp fragment covering part from genes *SMb20450* and *SMb20451*. All transposants showed a band size bigger than the wild type, suggesting the transposon was inserted in the StuI

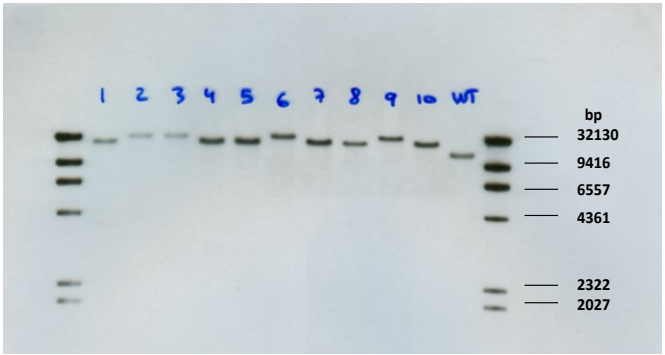


Figure 21. Southern Blot of transposants selected from a general mutagenesis showing a CR^+ phenotype (lanes 1-10). Genomic DNA was digested with StuI, which produces a fragment containing the entire *SMb20447-SMb20452* region (12,062 Kb, WT lane). The probe used binds between genes *SMb20450* and *SMb20451*. Lanes at both extremes correspond to lambda DNA/HindIII marker.

fragment (17.8 kb expected band). However, colonies 2, 3, 6 and 9 displayed an even bigger band. We deduced that in these colonies the entire vector along with the Tn*5* had been integrated, making the band look bigger than it should. In addition, the DNA of some of the colonies were sent for sequencing, but only in three of them the reaction worked well. In colonies 5, 7 and 8 the transposon was inserted in genes *SMb20449* and *SMb20450*.

Regarding this result, we cannot discard the existence of other genes elsewhere in the Sm8530 genome implicated in the regulation of the ML β-glucan, but it is significant that all the CR^+ transposants had the mutation in the same StuI fragment. Thus, we suggest that the operon *SMb20447-SMb20452* together with ExpR are the principal components of this regulatory pathway.

.

SMb20447 is the last element in the signal transduction system conformed by the operon and its activity is ultimately repressed by SMb20452.

The results obtained so far pointed to the *SMb20447-SMb20452* operon as a probable signal transduction system that regulates the ML β-glucan synthesis. With the aim of unravel the hierarchy of the six components along the signal cascade, we constructed a battery of single, double, and triple mutant strains, that were subjected to crossed complementation analysis (Table 6). All the strains that produce the β-glucan (Δ*SMb20449*, Δ*SMb20450* and Δ*SMb20452*) reverted to CF^- and CR^- phenotypes when *SMb20447* was not present. Similarly, the overexpression of *SMb20451* (pJL2112) did not produce the ML β-glucan in a ΔSMb20447 strain. Most interestingly, the expression of *SMb20452* (pJL2113) could neither complement a strain lacking both *SMb20450* and *SMb20452*, but it did restore the wild type CF^- phenotype in a strain lacking the last three genes of the operon (*SMb20450*, *SMb20451* and *SMb20452*). Thus, it seems that the ability of *SMb20452* to restore the wild type phenotype is determined by the presence or absence of *SMb20451*. This is consistent to what we observed in the Δ*SMb20451* strain, which could only be complemented by *SMb20452* (if we assumed the polar effect of the mutation in *SMb20451*, we could consider this mutant like a double mutant *SMb20451-SMb20452*).

As a summary, this genetic analysis has focused in the four genes of the operon that presented the most pronounced CF^+ and mucoid phenotypes. SMb20447 is the last element that transduces the signal to BgsA, whereas the position of SMb20452 in the signal

cascade should be immediately upstream of SMb20447, acting as a negatively regulator. Interestingly, this negative regulation of SMb20452 over SMb20447 can be observed only under certain conditions: when both SMb20450 and SMb20451 are either present/absent, or only if SMb20451 is absent.

Table 6. Genetic analysis and complementation tests of different mutants from the operon based on their CF phenotype. (++) denotes a more intense fluorescence than (+).

Strain	Presence/Absence of the native gene in pSymB						Gene expressed *in trans*	CF phenotype
	SMb20447	*SMb20448*	*SMb20449*	*SMb20450*	*SMb20451*	*SMb20452*		
IBR502	✓	✓	×	✓	✓	✓	-	+
IBR503	✓	✓	✓	×	✓	✓	*SMb20450* *SMb20452*	- ++
IBR505	✓	✓	✓	✓	✓	×	*SMb20450* *SMb20452*	++ -
IBR518	×	✓	×	✓	✓	✓	- *SMb20447*	- ++
IBR519	×	✓	✓	×	✓	✓	- *SMb20447*	- ++
IBR520	×	✓	✓	✓	✓	×	- *SMb20447*	- ++
Sm8530	✓	✓	✓	✓	✓	✓	*SMb20451*	++
IBR500	×	✓	✓	✓	✓	✓	*SMb20451*	-
IBR504	✓	✓	✓	✓	×	✓	*SMb20451* *SMb20452*	+ -
IBR521	✓	✓	✓	×	✓	×	- *SMb20452*	++ ++
IBR522	✓	✓	✓	×	×	×	- *SMb20452*	++ -

(*) denotes the possibility of considering the Δ*SMb20451* mutant as a double mutant *SMb20451*/*SMb20452* due to the putative polar effect of the deletion in *SMb20451*.

SMb20450, SMb20451 and SMb20452 present domains similar to the proteins involved in the regulation of the stress response and sporulation in *Bacillus subtilis.*

To further deepen in the putative function of the proteins coded within the operon, we checked the presence of conserved domains in their amino acid sequences (Marchler-Bauer *et al.*, 2015). SMb20448 did not show any putative function or specific domain. SMb20449 is annotated as a type 11 methyltransferase with specific SAM domain (S-

AdenosylMethionine-dependent methyltransferases). SMb20450 is a regulatory protein with Cache, HAMP and PP2C domains. Cache domains are generally extracellular motifs that bind small molecules and convert the signal into different outputs depending on their intracellular effector domains (Anantharaman & Aravind, 2000). HAMP domains received that designation because they are found in histidine kinases, adenylyl cyclases, methyl-binding proteins and phosphatases. They participate in signal transduction events (Aravind & Ponting, 1999). PP2C are Mn^{2+} or Mg^{2+} dependent serine/threonine phosphatases. Continuing with the remaining proteins, SMb20451 is predicted to bind ATP and presents an ATPase-like domain characteristic from kinases. Finally, SMb20452 displays a STAS domain typical from anti-sigma factor antagonists.

The well characterized *Bacillus subtilis* have two sets of proteins with PP2C, ATPase-kinase and STAS domains named RsbU/SpoIIE, RsbW/SpoIIAB and RsbV/SpoIIAA that conform each a system controlling the activity of sigma factors σ^B and σ^F respectively. Ultimately σ^F activates the genes involved in the sporulation process, whereas σ^B controls the transcription of those genes belonging to the stress response. Under a non-stress situation, the sigma factors are arrested by their cognate anti-sigma factors that possess kinase activity (SpoIIAB or RsbW). This complex can be disrupted by the binding of a non-phosphorylated form of the anti-sigma factor antagonists (STAS proteins SpoIIAA or RsbV), so the sigma factor is released and can initiate the transcription of the corresponding genes. Eventually SpoIIAB and RsbW phosphorylate the STAS proteins, provoking their removal, and allowing the re-binding of the anti-sigma factors to σ^B and σ^F (Hecker *et al.*, 2007, Yudkin & Clarkson, 2005). This sort of regulatory complexes conformed by a phosphatase, a kinase and a STAS proteins have been described in other gram positive and negative bacteria such as *Mycobacterium tuberculosis*, *Chlamydia trachomatis*, *Vibrio fischeri*, or *Bordetella sp* (Kozak *et al.*, 2005, Hua *et al.*, 2006, Morris & Visick, 2013b). They belong to the so-called "partner switching" systems, which are based on protein-protein interactions and reversible phosphorylation reactions (Francez-Charlot *et al.*, 2015).

The serine kinases and STAS proteins from these systems presents conserved domains and amino acid residues that have been demonstrated to be critical for the phosphorylation events (Sharma *et al.*, 2011, Dutta & Inouye, 2000). Thus, we performed a multiple sequence alignment of SMb20451 and SMb20452 with other well studied "partner switching" proteins from different organisms (Figure 22). The STAS proteins showed a conserved serine around position 60, which was also present in SMb20452 (S61). Several

studies have demonstrated that this residue is the phosphorylation target (Yang *et al.*, 1996, Najafi *et al.*, 1995). On the other hand, kinases from partner switching systems are typical constituted by an ATP-binding Bergerat fold with four conserved motifs called N, G1, G2 and G3 boxes involved in nucleotide binding. Among these motifs, several specifics residues have been demonstrated to be necessary for ATP binding: the asparagine residue from the N box, and the aspartate and glycine residues from G1 box, and the two glycines from the G2 box (Dutta & Inouye, 2000). All of them are also present in SMb20451. Considering the results from the genetic/complementation analysis, and the homology that some of the elements from the operon share with partner switching systems, we predicted that the phosphorylation state of SMb20452 is determinant for the activation/inhibition of the ML β-glucan. SMb20450 and SMb20451 would be the phosphatase and kinase partners regulating the STAS protein.

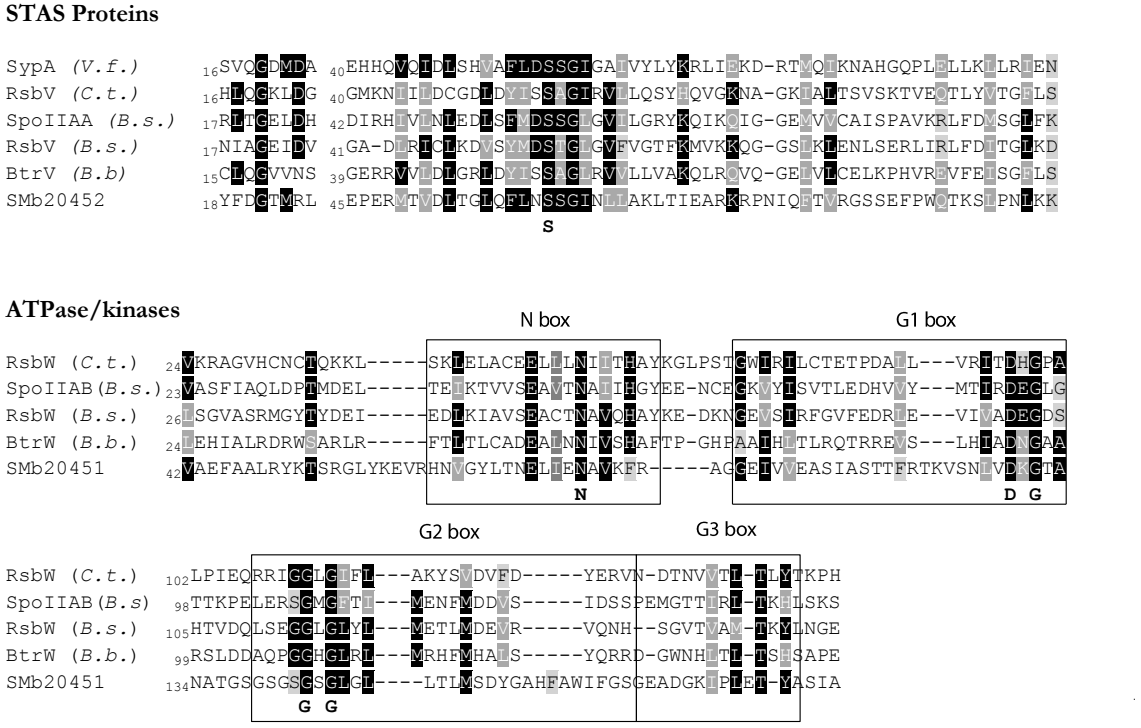


Figure 22. Multiple sequence alignment of STAS proteins and ATPase/kinases from partner switching systems in different organisms compared to SMb20452 and SMb20451 respectively. White letters over black backgrounds indicate identical amino acids; white letters over grey background denote conservative substitutions; black letters over grey backgrounds indicate semi-conservative substitutions. The different organisms are indicated between brackets: *Chlamidya trachomatis* (*C.t.*)*; Bacillus subtilis* (*B.s.*)*; Bordetella bronchiseptica* (*B.b.*)*; Vibrio fischeri* (*V.f.*).

The non-phosphorylated form of SMb20452 inhibits the ML β-glucan production.

To verify whether phosphorylation reactions over SMb20452 were important for the β-glucan regulation, we substituted the serine at position 61 with an alanine (SMb20452^{S61A}), which mimics a permanent non-phosphorylated state (Yang *et al.*, 1996), and constructed a new plasmid for overexpression purposes (pJL2113S61A). Interestingly, the introduction of pJL2113S61A into a strain defective in the alleged phosphatase (Δ*SMb20450*) restored the CF^{-} wild type phenotype (Supplementary Table 2), in contrast to what we observed when we introduced a wild type copy of *SMb20452* in the same genetic background (Table 6). According to our hypothesis, in the Δ*SMb20450* mutant background, all SMb20452 would be phosphorylated since the kinase SMb20451 is the only partner acting. Thus, only by introducing a permanent non-phosphorylated form of SMb20452 we would be able to restore the wild type phenotype.

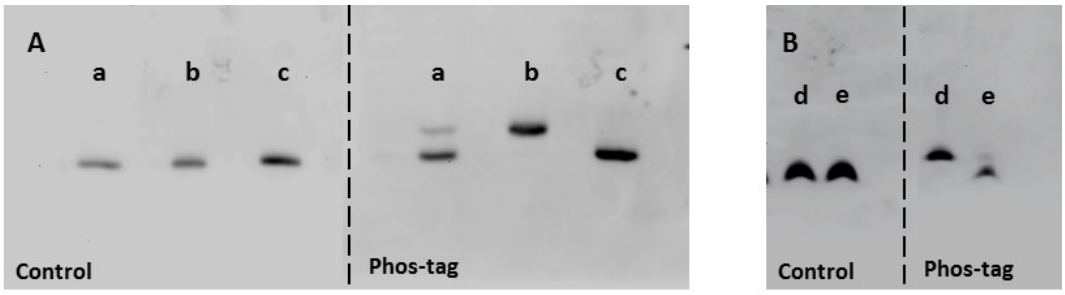


Figure 23. *In vivo* phosphorylation assay of c-Myc tagged SMb20452 in different genetic backgrounds: **(a)**Sm8530, **(b)** Δ*SMb20450* strain, **(c)** Δ*SMb20451* strain, **(d)** Δ*SMb20450* pFAJ1708, **(e)** Δ*SMb20450* pJL2111. Phosphorylated and non-phosphorylated forms correspond to the upper and lower bands, respectively.

Additionally, we constructed a c-Myc tagged version of SMb20452 and performed *in-vivo* phosphorylation assays under different genetics backgrounds: wild type, Δ*SMb20450* and Δ*SMb20451* mutant strains. In this type of assays, phosphorylated or non-phosphorylated versions of the same protein can be distinguished thanks to their different migration rates in acrylamide gels with a reactive (Phos-tag™) that specifically binds to phosphorylated ions (Kinoshita *et al.*, 2015). Figure 23A shows that SMb20452 coexists in both states in a wild type background. The Δ*SMb20450* strain, however, only presents the upper band corresponding to the phosphorylated form, whereas in the Δ*SMb20451* strain a single lower band (non-phosphorylated) is detected. The introduction of a copy of *SMb20450* (pJL2111) in the Δ*SMb20450* mutant strain restored the wild type phenotype,

which supports that SMb20450 is indeed acting as a phosphatase (Figure 23B). This is consistent with the results obtained from the crossed complementation analysis (Table 6), where SMb20452 is able to restore the wild type phenotype under certain conditions, namely when SMb20451 (the kinase) is not present. Taken together our findings indicate that the non-phosphorylated form of SMb20452 inhibits the DGC activity of SMb20447 and in last term the ML β-glucan production.

Construction of tagged proteins to study interactions between SMb20447, SMb20451 and SMb20452.

In the partner switching systems that have been described so far, protein-protein interactions in addition to reversible phosphorylation events have proved to be very important. In our case, we speculated that SMb20452 could physically interact, at least transiently, with the phosphatase SMb20450 and the kinase SMb20451 just simply to either receive or get rid of the phosphate group. However, the type of inhibition that SMb20452 exerts over the DGC SMb20447 remained unclear. To confirm or discard the existence of physical interactions, we constructed tagged versions at the N-terminus of the putative soluble partners of the system (StrepII-SMb20447, HA-SMb20451 and c-Myc-SMb20452) with the aim of performing in-vitro pull down assays. These versions were expressed from an arabinose inducible pMLBAD vector and produced the expected phenotypes when introduced in the required genetic background, suggesting that the proteins were functional. Unfortunately, we did not succeed to obtain the sufficient amount of soluble StrepII-SMb20447 and c-Myc-SMb20452 necessary for the pull down assays.

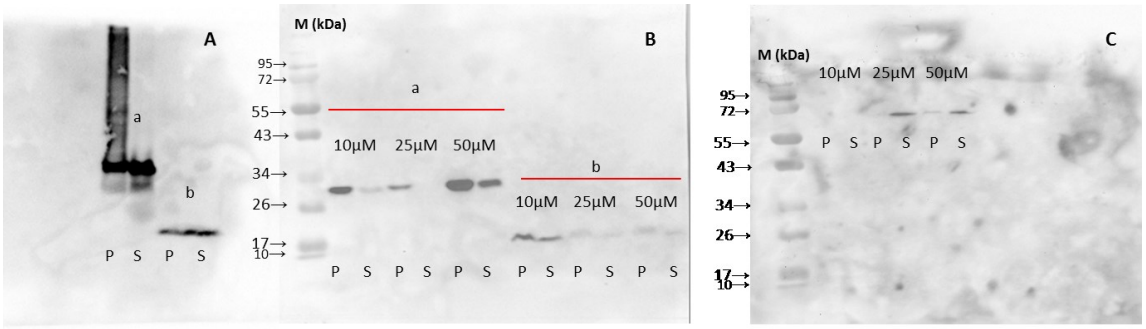


Figure 24. Western blots of HA and c-Myc tagged versions of SMb20447, SMb20451 and SMb20452 expressed from cumate inducible pQF derivatives in Sm8530. Letters P and S indicate the pellet or supernatant fraction of the sample. **A.** HA tagged versions of SMb20451 **(a)** and SMb20452 **(b)**. The expected sizes are 21.29 kDa and 14.21 respectively. kDa. **B.** c-Myc tagged versions of SMb20451 **(a)** and SMb20452 **(b)** using different concentrations of cumate. The expected sizes are 21.52 kDa and 14.35 kDa respectively. **C.** c-Myc tagged versions of SMb20447 using different cumate concentrations. The expected size is 63.91 kDa.

We decided then to change the tags to the C-terminus of the proteins and also used other expression vectors based on the pQF plasmid (see Materials and Methods). This time we constructed nine plasmids derivatives (Table 2), in order to have each protein with the three different tags and by this way dispose of a wide range of possible combinations. The new tagged versions of SMb20451 and SMb20452 produced the expected phenotypes when overexpressed in the required background, suggesting that the proteins were functional; they also yield a good amount of the soluble fraction in western blot analysis. SMb20447, however, could only be detected with the c-Myc tag, and the protein produced was significantly lower. Further experiments for expression optimization from pQF derived plasmids will be needed, in order to be able to perform the pull-down assays. Figure 24 shows western blots with anti HA and c-Myc antibodies for detecting tagged versions ofSMb20447, SMb20451 and SMb20452.

Role of the ML β-glucan in the alfalfa rhizosphere.

As mentioned in the introduction section, the relevance of rhizobial exopolysaccharides in the rhizosphere has been well proved. Thus, we wondered whether the ML β-glucan was also involved in any of these processes. To test the nodulation capacity of the ML β-glucan, we inoculated alfalfa seedlings with a Δ*SMb20450* strain with two additional mutations that blocked the production of both EPS I and EPS II (strain IBR508). After four weeks all the nodules that had been developed were white, suggesting that bacteria could not be able to reach the nodule and start the fixation (data not shown).

Regarding the attachment to alfalfa roots, we confirmed the results obtained by Perez-Mendoza *et al.* (2015). A mutant defective in the ML β-glucan (*Smb20391::Nm*) was progressively displaced by the wild type (Figure 25A). Surprisingly, the strain Δ*SMb20450*, which constitutively produces the β-glucan, was also displaced by the wild type although with a different pattern. The Δ*SMb20450* population dropped drastically at 24h from the inoculation (80% Sm8530 versus 20% Δ*SMb20450*) but then it was maintained at this proportion (Figure 25B). Previously, Rinaudi and González (2009) demonstrated the importance of the EPS II (specially the low molecular weight form) at the initial attachment to the root. Thus, we wondered whether the EPS II deficiency in the strain Δ*SMb20450* could be self-defeating. To overcome this drawback, we performed a new assay using EPS II defective variants of the wild type and the Δ*SMb20450* strain. As shown in Figure 25C,

this time both strains were equally present in the roots through all the experiment. Finally, we wanted to analyse the effect of SMb20447, since it seems to constitute the last step in the signal transduction pathway that regulates the ML β-glucan production. No alterations were observed (Figure 25D), suggesting that the strain was able to produce the ML β-glucan despite the absence of SMb20447.

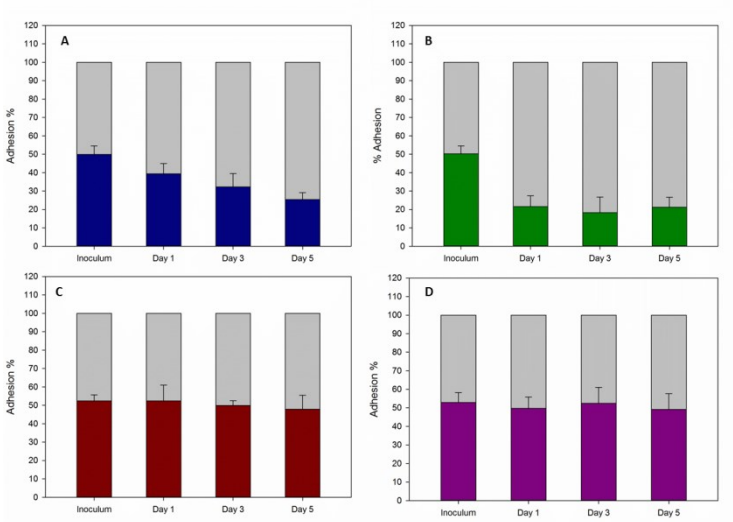


Figure 25. Competitive attachment to alfalfa roots of different strain pairs. The roots were inoculated with an equal mixture (10^5 cells) of both strains (first bar on the graphs), and then attached bacteria were recovered at 1 day, 3 days, and 5 days. Data is expressed as a percentage of mutant CFUs out of the total. Grey bars stands for the wild type strain. **A.** Sm8530 Vs *SMb20391::Nm*. **B.** Sm8530 Vs Δ*SMb20450*. **C.** Sm8530 EPS II⁻ Vs Δ*SMb20450* EPS II⁻ .**D.** Sm8530 Vs *Smb20447::Gn*

Taken together, these results suggest that the ML β-glucan does not substitute EPS I or EPS II in the nodule invasion process. However, it is clear that the ML β-glucan plays an important role in the previous attachment to the roots, since a strain lacking the synthase (SMb20391) is displaced by the wild type. Interestingly, a strain that constitutively produces the ML β-glucan does not get any advantage over the wild type, suggesting that its effectiveness could be rather at low levels or specific niches. Coinciding with the findings from Perez-Mendoza *et al.* (2015), the ML β-glucan is neither important for the attachment to abiotic surfaces like polystyrene, as the Δ*SMb20450* strain behaved like the wild type strain Sm1021 (Figure 13A). On the other hand, we should highlight the importance of the EPS II during the first 24h of the assay, since the Δ*SMb20450* strain is clearly displaced from the beginning. Considering the patterns of Figures 24A and 24B, it seems that EPS II is more important at early stages, whereas the ML β-glucan might be necessary through the whole attachment process. Finally, the absence of SMb20447 does not produce any

deleterious effect. As outlined above, maybe the production of the ML β-glucan at basal levels is sufficient for the attachment, or there is another DGC that can substitute SMb20447.

Global transcriptomic analysis revealed that SMb20450 potentially regulates two other EPS and several genes involved in stress response, rhizobactin synthesis, denitrification and cation uptake.

The fact that SMb20450 is annotated as a transmembrane regulatory protein made us consider it could also be regulating other functions apart from the ML β-glucan synthesis. Thus, we performed a microarray analysis comparing Sm8530 versus a Δ*SMb20450* strain. Cultures were grown in minimal medium and the transcriptome were analysed at early stationary phase. A total of 127 genes were differentially regulated, and nearly 70% were induced in the mutant. The complete set of genes differentially regulated is listed in Supplementary Tables 3 and 4.

Down-regulated genes in the Δ*SMb20450* strain.

Thirty nine genes appeared to be down-regulated in the strain Δ*SMb20450*, the majority of them located in the pSymB megaplasmid (21), and the rest distributed among the pSymA (8) and the chromosome (10). Among the pSymB genes, it is worthy to distinguish the presence of ten genes that belong to the three operons responsible for the synthesis and secretion of the EPS II (*wgeABCH*, *wgdB*, and *wgaABDFJ*). This is consistent with the EPS II defective phenotypes observed in this strain. There are also genes associated to nitrogen regulation, metabolism, and assimilation (*SMc03806, SMc03807, SMb20984, SMb20745,* and *SMa0585*), translation of proteins (*SMc01313*), and general DNA/ARN metabolism. A general cluster analysis based on the GO annotation revealed that the main biological and molecular functions repressed in the Δ*SMb20450* strain are related to nucleic acid and nitrogen metabolism, and polysaccharide biosynthesis (Figure 26).

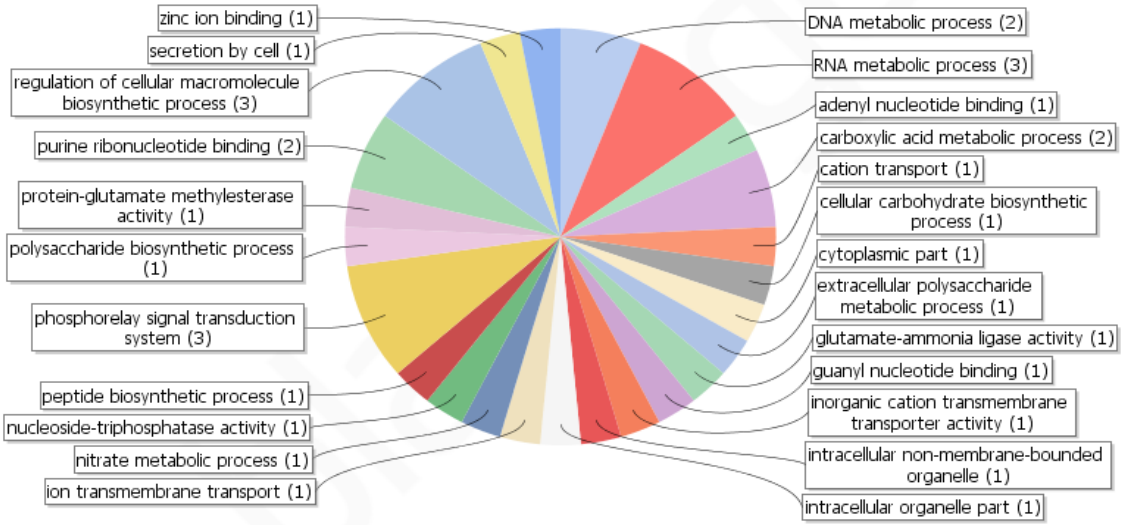


Figure 26. Blast2GO cluster analysis showing the general biological and molecular functions that are repressed in the strain Δ*SMb20450*.

Up-regulated genes in the Δ*SMb20450* strain.

The number of induced genes (88) is significantly high compared to the repressed ones. More than 50% belong to the pSymA megaplasmid (49 genes), followed by chromosome genes (26), and pSymB genes (13). We found genes related to the rhizobactin 1021 synthesis and secretion (*rhbABCEF*, *rhtA*, *rhtX*, *SMa2339*) and to the uptake of other siderophores (*hmuVTS, SMc02726*). There were also six genes coding for proteins of the Universal Stress Response (Usr) family (*SMa1095*, *SMa1100*, *SMa1147*, *SMa1149*, *SMa1158* and *SMa1231*), and a set of genes involved in nitrate respiration and denitrification (*nosZDFYL*, *napB*, *nirV*, *NnrU*). Genes *fixQ1* and *fixP1* were as well in the list of induced genes. They form part of the operon *fixNOPQ* coding for proteins of the cbb3 oxidase, which operates in free-living and micro-oxic conditions. Regarding the pSymB megaplasmid, we found two transcriptional regulators belonging to the LuxR and AraC families (*SMb20162*, *SMb20344*) and, very interestingly, *SMb20460*, the other gene annotated as a cellulose synthase, together with the three immediately down-stream genes *SMb20461, SMb20462* and *SMb20463*.

Likewise with the repressed genes, we performed a cluster analysis based of the GO available annotations to have a general idea of the most common biological processes and molecular functions induced in the Δ*SMb20450* strain. As shown in Figure 27, it seems the majority of the genes are involved in siderophore biosynthesis and uptake, amino acid metabolism, cellular respiration, and transmembrane transport processes. In relation to

this calcium transport, we obtained very preliminary results showing that the strains Δ*SMb20450* and Δ*SMb20452* are able to reach higher optical densities than the wild type when grown in liquid TY medium without $CaCl_2$ (Supplementary Figure 1).

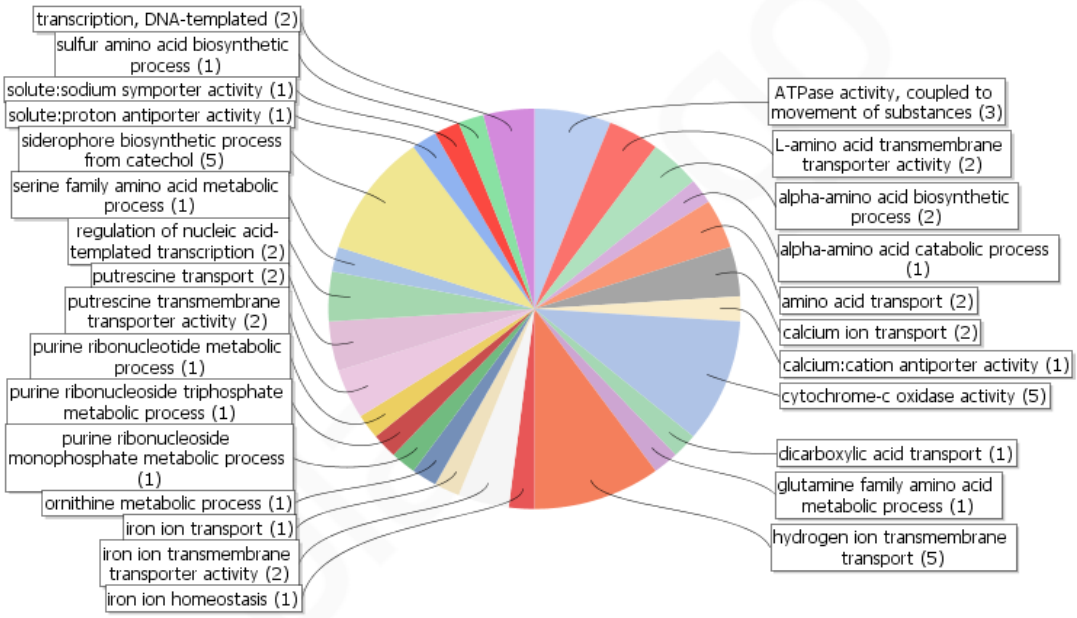


Figure 27. Blast2GO cluster analysis of biological processes and molecular functions up-regulated in the strain Δ*SMb20450*.

The operon *SMb20447-SMb20452* is conserved among rhizobiales.

It has been shown that the *bgsAB* operon responsible for the synthesis of the ML β-glucan is present in the *Sinorhizobium*, *Rhizobium*, *Agrobacterium* and *Methylobacterium* genera within the Rhizobiales (Perez-Mendoza *et al.*, 2015). The *SMb20447-SMb20452* operon is also present in Rhizobiales, but only in the Rhizobiaceae and not in the Methylobacteriaceae (Figure 28). It is also present in *Mesorhizobium loti* MAFF303099 although this bacterium does not present genes homologous to *bgsA* and *bgsB*, needed for the synthesis of the polysaccharide. Interestingly, in this bacterium the operon is truncated

by three small and opposite oriented ORFs located between the DGC and the rest of the genes in the operon. Outside the Rhizobiales, similar gene clusters, with at least a partial version with the DGC, the phosphatase, the kinase and the protein bearing the STAS domain is found in *Rhodospirillum*, *Shorangium*, *Thioalkalivibrio* and *Magnetospirillum* genera (Figure 28). In some cases the kinases and STAS proteins are either duplicated or fusioned.

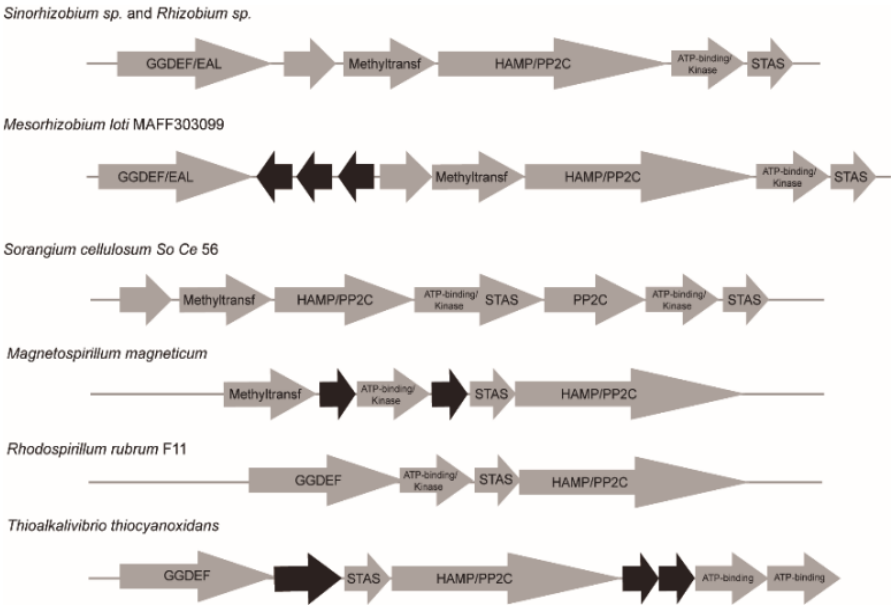


Figure 28. Homology of the *SMb20447-SMb20452* operon in different rhizobia and non-rhizobia species.

DISCUSSION

In this study, a screening of mutants affected in genes coding for putative c-di-GMP metabolism proteins led us to identify a pSymB operon involved in the regulation of the recently described ML β-glucan in *S. meliloti* (Perez-Mendoza *et al.*, 2015). Our investigations focused on SMb20447, a predicted cytosolic protein that presents both the consensus GG(D/E)EF and EAL domains with all the critical amino acids necessary for their corresponding activities (Romling *et al.*, 2013, Bobrov *et al.*, 2011, Krogh *et al.*, 2001). The only exception is the presence of an alanine located at the first position of the DGC motif. The putative DGCs displaying either an AGDEF or SGDEF motifs represents less than 3% out of the GGDEF (PF00990) domain listed in the last release of the Pfam database (Finn *et al.*, 2014). However, there is an increasing number that have proved to be active (Hunter *et al.*, 2014, Perez-Mendoza *et al.*, 2011), suggesting that the consensus GG(D/E)EF motif is not as restricted as it was believed. In our case, we have found several evidences that confirm SMb20447 is active as a DGC at least when overexpressed: reduction in swimming motility, increase in ML β-glucan production, and high intracellular c-di-GMP levels. The substitution of the glutamic acid from the AGDEF motif with an alanine (AGDAF) abolished the CF^{+} phenotype above mentioned and almost restored the wild type swimming motility, supporting the role of SMb20447 as a DGC.

Interestingly, the overexpression of this AGDAF version in the Δ*SMb20450* genetic background provokes the reversion to the wild type phenotype regarding CF^{-} and mucoid aspect, in contrast to the overexpression of the original *SMb20447* in the same background. This preliminary result gives rise to new hypotheses based on the dual activity of SMb20447 as a DGC and PDE. Only a few samples of truly bifunctional DGCs/PDEs have been described so far : *Rhodobacter Sphaeroides* BphG1, *Vibrio parahaemolyticusK* ScrC, *Legionella pneumophila* Lpl0329, *Pseudomonas putida* PP2258, *Mycobacterium smegmatis* MSDGC-1 and *M. tuberculosis* MtbDGC (Gupta *et al.*, 2010, Tarutina *et al.*, 2006, Ferreira *et al.*, 2008, Levet-Paulo *et al.*, 2011, Osterberg *et al.*, 2013, Kumar & Chatterji, 2008). Most recently, Gao *et al.* (2014) have described two proteins in *Rhizobium etli,* CdgA and CdgB, with both GGDEF and EAL active motifs, at least when the domains are single overexpressed in *E. coli*. More common are the cases where the EAL domain has lost its catalytic activity but still binds c-di-GMP or participates in protein-protein or protein-ARN interactions (Romling *et al.*, 2013). Further analysis concerning point mutations, complementation tests, and intracellular c-di-GMP measurements will be needed to confirm these hypotheses.

RT-PCR analysis revealed that *SMb20447* is the first gene of an operon comprising six genes, from *SMb20447* to *SMb20452*. A global mapping of transcription start sites performed by Schluter *et al.* (2013) supports this result and also predicts the presence of additional internal promoter regions in the operon, especially in *SMb20451*. This would explain the behavior of mutant Δ*SMb20451*, which presented an intermediate CF and mucoid phenotype between the wild type and the Δ*SMb20452* mutant, and could only be complemented by the introduction *in trans* of *SMb20452*. Another possibility would be the occurrence of sequence mistakes during the homologous recombination in the Δ*SMb20451* mutant generation that affected the expression of *SMb20452*. It will be interesting to identify empirically these putative internal promoters and investigate how (or by whom) they are induced or repressed.

At the transcriptional level, it has been shown that the ML β-glucan production is dependent on quorum sensing, as the transcription of the operon responsible for its synthesis, *bgsBA*, is activated by the Sin/ExpR regulon (Perez-Mendoza *et al.*, 2015). This finding was confirmed also by our results. However, this is not the case for the *SMb20447-SMb20452* operon, which present similar levels of expression in the ExpR$^+$ and ExpR$^-$ strains in the conditions tested, suggesting it forms an independent regulatory network for controlling the synthesis of the polysaccharide. Interestingly, a small increase in the transcriptional levels of a *bgsA*::*lacZ* fusion was found in a Δ*SMb20450* background. A similar increase was found by q-PCR experiments in strains overexpressing *pleD**, although it was neither statistically significant nor confirmed by *bgsA*::*gusA* fusion (Perez-Mendoza *et al.*, 2015). The microarray data from the Δ*SMb20450* strain did not show any induction of *bgsA* either, probably because the small increment detected with the *lacZ* fusions is below the "at least two-fold" threshold stablished for the microarray data treatment. Thus, although the main regulatory role that c-di-GMP has over the synthesis of the ML β-glucan is posttranscriptional, it has to be considered that it could also play a minor role at the transcriptional level, adjusting *bgsBA* expression to specific conditions. c-di-GMP posttranslational regulation is common among bacterial cellulose synthases, where the second messenger induces the enzyme activity by binding to its PilZ domain (Romling *et al.*, 2013). Interestingly, BgsA does not present a PilZ domain, but there are strong evidences showing its capacity to bind c-di-GMP (Perez-Mendoza *et al.*, 2015). As mentioned in the introduction, the PilZ domain has been for years the only c-di-GMP effector described in polysaccharide synthases and many other types of proteins. But there are recent descriptions of proteins involved in EPS synthesis like PelD from *P. aeruginosa* or

BcsE from *S. enterica*, that bind c-di-GMP through a domain very similar to the inhibitory site (I-site) of the DGCs (Lee *et al.*, 2007, Fang *et al.*, 2014). There is another case in *E. coli*, where c-di-GMP binding is necessary to keep stable the PgaCD membrane complex involved in the poly-β-1,6-N-acetylglucosamine (Poly-GlcNac) synthesis and export (Steiner *et al.*, 2013).

We also determined that SMb20447 is not likely regulated transcriptionally by the other genes in the operon, as the transcriptional levels of this gene do not change in a mutant lacking *SMb20450*. So probably, the *SMb20447-SMb20452* operon is constitutively expressed, regulating c-di-GMP associated phenotypes through the activity of SMb20447. The results obtained from the general mutagenesis suggest that the *SMb20447-SMb20452* operon is the main regulator of the ML β-glucan together with ExpR. The co-occurrence of the operon in most of the Rhizobiaceae bacteria that also bear the *bgsBA* operon suggests that c-di-GMP levels triggering the synthesis of the ML β-glucan are regulated in a similar way. It will be interesting to analyze whether high levels of c-di-GMP also promotes the synthesis of the same polysaccharide in *Methylobacterium*, which does not present the regulatory operon but it does have the *bgsBA* one. On the other hand, the effect that this operon has over the levels of a crucial second messenger may have important consequences in several traits that have not been considered in this work. Similar operon architectures have been found in other bacteria from very different environments: *Rhodospirillum rubrum* and *Magnetospirillum magneticum* are α-proteobacteria. The former is a photosynthetic nitrogen-fixing bacterium, and the latter is a magnetotactic bacterium. *Thioalkalivibrio thiocyanoxidans* (γ-proteobacteria) is an haloalkaliphilic, autotrophic, sulfur-oxidizing bacterium, and *Sorangium cellulosum* (δ-proteobacteria) is a myxobacteria known by its capacity of producing a wide range of secondary metabolits (Schneiker *et al.*, 2007, Sorokin *et al.*, 2002, Wang *et al.*, 2013, Wang & Norén, 2006).

Deletion mutants and overexpression assays revealed that the six proteins coded by the operon participate to a greater or lesser extent in the synthesis of the ML β-glucan, acting as positive or negative regulators. Homology and conserved domains analysis showed that at least three of them likely form a partner switching mechanism that regulates the activity of the DGC SMb20447. Similar type of partner switching systems are well known in *Bacillus subtilis* and other related bacteria as regulators of the stress and sporulation responses through its interaction with the corresponding sigma factors (Yudkin & Clarkson, 2005, Hecker *et al.*, 2007). They are characterized by the presence of a protein

with a PP2C-like domain, an anti-sigma factor with serine-kinase activity and an anti-sigma factor antagonist with a STAS domain (Sharma *et al.*, 2011, Hecker *et al.*, 2007). In other non-gram positive bacteria this type of partner switching mechanism has been described in *Chlamydia* (Hua *et al.*, 2006) and the proteobacteria *Bordetella* and *Vibrio fischeri* (Mattoo *et al.*, 2004, Visick, 2009). In *Bordetella*, the mechanism regulates the type III secretion system (TTSS) through a complex regulatory cascade involving protein interaction between the partner switchers themselves and also with still uncharacterized partners that transmit the signal to the TTS apparatus (Kozak *et al.*, 2005, Mattoo *et al.*, 2004). On the other hand, the *V. fischeri* SypE/SypA partner switching mechanism regulates the biofilm formation on the surface of the symbiotic light organ in the squid *Euprymna scolopes* (Morris & Visick, 2013b, Thompson *et al.*, 2015). This system is somehow unusual as there is only one protein, SypE, displaying both the kinase and the phosphatase activities that act over the anti-sigma factor antagonist like protein SypA. Both domains are separated by a central REC domain that present an aspartic acid residue whose phosphorylation status controls the opposing activities of the protein (Morris *et al.*, 2011). Unphosphorylated SypA promotes biofilm formation through a still unknown pathway that seems to act below the level of other *syp* genes transcription (Morris & Visick, 2013a, Morris & Visick, 2013b). Given the similarity of our system with the partner switching systems above mentioned, and taking into account their nomenclature, we have proposed to rename the *SMb20447-SMb20452* operon as *bgrRSTUWV* (for β-glucan regulation). BgrR stands for the DGC SMb20447; BgrU is the PP2C phosphatase SMb20450; BgrW the serine kinase SMb20451, and BgrV the STAS protein SMb20452 (Figure 29).

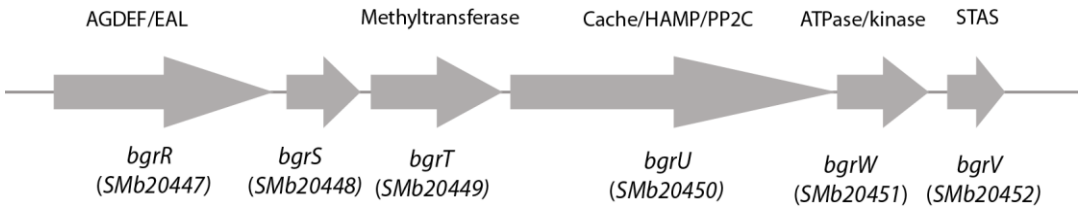


Figure 29. Genetic map of the *bgr* operon, showing the originally and the new nomenclatures, as well as the domains presented by their corresponding proteins.

The characterization of the *bgr* operon in *S. meliloti* is, to our knowledge, the first report of a partner switching mechanism that likely acts upon the activity of a DGC (BgrR), regulating then the c-di-GMP dependent synthesis of the ML β-glucan (Perez-Mendoza *et*

al., 2015). Deletion in the genes coding for the protein with a phosphatase domain, *bgrU*, or the STAS domain, *bgrV*, promotes the synthesis of the polysaccharide, indicating their negative regulatory activity. On the other hand, the overexpression of the gene coding for a putative serine kinase protein, *bgrW*, has the same phenotype, showing its opposite role. Among the proteins of the operon, BgrS (SMb20448) and BgrT (SMb20449) have the least effect on ML β-glucan dependent calcofluor fluorescence, although they are present in all the Rhizobial strains that harbor the operon in their genomes. The antagonistic effect of overexpression and mutation of these two genes indicate an opposite role in the regulation of c-di-GMP levels. The homology of BgrT with methyl transferases suggests that it might play a role in the signal transduction cascade through BgrU. BgrU possess a HAMP domain which is found in histidine kinases, adenylyl cyclases, methyl-accepting chemotaxis proteins and some phosphatases (Hazelbauer & Lai, 2010). This domain may be subjected to conformational changes responding to environmental stimuli and mediated by methylation of specific residues within the protein (Zhou *et al.*, 2009). In addition, BgrU displays an extracytoplasmic Cache domain that likely acts by sensing and binding small signal molecules (Anantharaman & Aravind, 2000).

Double/triple mutants and complementation analysis have allowed us to establish the hierarchy of the system. Strains lacking the DGC BgrR do not produce the ML β-glucan unless a source of c-di-GMP, such as PleD*, provides the second messenger to the cells and then, the polysaccharide production is not regulated through the system. The DGC activity of BgrR seems to be regulated by the partner switching formed by the other proteins in the operon. Specifically, BgrV alone could act as a repressor of DGC activity of BgrR, because a strain carrying a deletion in *bgrUWV* presents a ML β-glucan positive phenotype that can be complemented by the introduction *bgrV* alone. Interestingly, the presence of BgrV does not revert the wild type phenotype in those mutants where only the phosphatase BgrU is absent. Thereby, the role of BgrU and BgrW should be to modify the phosphorylation status of BgrV, leading to a fine control of BgrR activity, as it has been demonstrated in the other similar systems (Hua *et al.*, 2006, Kozak *et al.*, 2005, Morris & Visick, 2013b, Sharma *et al.*, 2011). Amino acid sequences alignment of BgrV with other STAS proteins from partner switching systems revealed the presence of a conserved serine at position 61 which has been proved to be the phosphorylation target (Yang *et al.*, 1996). To further deepen in this hypothesis, we substituted this serine with an alanine residue that mimics a "permanent" non-phosphorylated status (Yang *et al.*, 1996). This version $BgrV^{S61A}$ was introduced in the strain Δ*bgrU* and caused the reversion to the wild type phenotype,

contrary to what was observed when a normal BgrV was expressed in the same genetic background. Furthermore, *in-vivo* phosphorylation assays with a c-Myc tagged version of BgrV under different genetic backgrounds revealed that BgrV is present in both states, phosphorylated and non-phosphorylated, in the wild type strain. In the Δ*bgrU* strain, where BgrW is present, BrgV appears only phosphorylated, whereas in the Δ*bgrW* strain, which has the phosphatase BgrU, we could only detect the non-phosphorylated form. The introduction of *bgrU* in *trans* in the Δ*bgrU* mutant strain promoted a switch of BgrV almost completely to the non-phosphorylated form, confirming the phosphatase activity of BgrU. Thus, it is clear that the phosphorylation status of BgrV is the key factor that determines the production of the ML β-glucan. The presence of non-phosphorylated form of BgrV is sufficient to inhibit or significantly reduce the activity of BgrR and then the production of the ML β-glucan. Based on the results above mentioned, we propose a model of how the partner switching system might work (Figure 30).

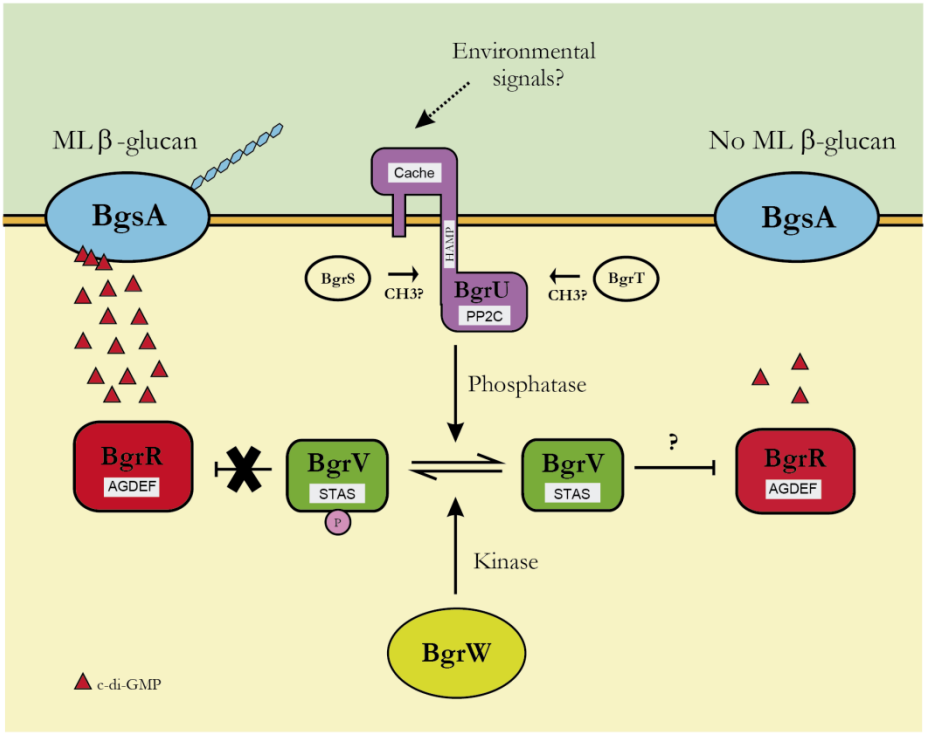


Figure 30. Putative partner switching system regulating the synthesis of the ML β-glucan through the activity of the DGC BgrR. The phosphorylation status of BgrV, which in turn is given by the balance between BgrU and BgrW activities, determinates the repression/activation of BgrR.

Although phosphorylation regulatory networks have been deeply studied in eukarya, the field is just emerging in bacteria. New technologies are facilitating quantitative

approaches to compare global phosphorylation patterns in different growth conditions or strains (Mijakovic & Macek, 2012). Recently, Liu *et al.* (2015) performed a site-specific Ser/Thr/Tyr phosphoproteome in *S. meliloti* and found 77 proteins, with serine constituting the majority of the phosphorylated residues (65%). These proteins are predicted to be involved in a wide range of molecular functions such as replication, translation, posttranslational modifications, transport and metabolism of amino acids, inorganic ions, carbohydrates, succinoglycan, stress adaptation, etc. Interestingly, BgrV appears among the 77 proteins, but not the three other proteins with a STAS domain found in the Sm1021 genome. Maybe the conditions tested were not favorable for the expression of these proteins, or they could not be detected due to technical limitations. Nevertheless, we suspect BgrV could be involved in other regulation processes, either at the transcriptional or posttranslational levels. It is not mentioned in the results section, but we also performed a microarray analysis comparing a Δ*bgrV* strain with the wild type Sm8530. Unfortunately, the quality of two out of the three replicates was not good enough to obtain any conclusive result, but we are planning to repeat this experiment in a near future.

The model proposed for the partner switching system has an important unsolved question: how is the activity of BgrR controlled by BgrV? It is common among DGCs to be activated or inhibited by phosphorylation (Romling *et al.*, 2013). However, BgrR does not present any additional domain that could be subjected to phosphorylation, as it is the case of the REC domain from PleD in *C. crescentus*. In addition, we have not found any references suggesting the ability of STAS proteins to phosphorylate other proteins. More likely, STAS proteins are involved in protein-protein interactions (Sharma *et al.*, 2011). Thus, the non-phosphorylated form of BgrV could bind to BgrR and prevent its DGC activity. Finally, there is another possibility related to substrate competition. The SpoIIAA STAS protein from *B.subtillis* has been proved to bind and hydrolyze GTP only in the non-phosphorylate state (Najafi *et al.*, 1996). It would be reasonable to think about the existence of some regulation based on the competition for GTP binding between the DGC and the STAS-domain protein. Further analysis including *in-vitro* pull-down assays will help to unravel whether there is a physical interaction between BgrR and BgrV.

Regarding the biological importance of the ML β-glucan, we confirmed what was recently described by Perez-Mendoza *et al.* (2015). A *bgsA* mutant is outcompeted by the wild type during the attachment to alfalfa root surfaces. However, if the ML β-glucan

regulatory pathway is permanently "on", it does not provide any advantage either, since both the mutant and wild type populations are equally present through time. The attachment pattern observed in strain Δ*bgrU* is due to its EPS II defective production. This result is consistent to the findings from Rinaudi and Gonzalez (2009), who demonstrated that the low-molecular weight form of the EPS II is determinant for biofilm formation and root colonization at early stages. Thereby, our results suggest that the EPS II is critical for the adhesion to the alfalfa roots during the first 24 hours, whereas the ML β-glucan seems to be necessary along the whole attachment process. However, we do not know if the presence of the ML β-glucan in small amounts is sufficient, or either is specifically synthesized at certain micro-niches of the root. Contrary to our findings, a strain that constitutively produces the ML β-glucan through the overexpression of the foreign DGC PleD* displays a better root attachment than the wild type (Perez-Mendoza *et al.*, 2015). The amount of ML β-glucan produced by strain Sm8530 pJBpleD* is remarkably high compared to the production in the Δ*bgrU* strain. Thus, we speculate that in the Sm8530 pJBpleD* strain the synthesis of the ML β-glucan does not correspond to a normal physiological response and is totally out-regulated, making the attachment to the root even more effective than with the EPS II. On the other hand, genetic/complementation analysis based on CF phenotypes have shown that DGC BgrR is the last step needed for the ML β-glucan production but it is not exclusive, since the addition of another ci-di-GMP source is able to substitute the absence of BgrR. In fact, a strain lacking BgrR seems to behave like the wild type regarding the attachment to the root, suggesting that either the ML β-glucan is permanently present at basal levels, or there is an alternative source of c-di-GMP. The environmental conditions in the root surface differ greatly from those in CF plates, and there might be process at the transcriptional/translational or post translational levels that we are not taken into account (Giddens *et al.*, 2007). For instance, there are cases of DGCs that are niche-specific, like PP4959 from *P. putida* 2440, whose promoter is induced by corn plant root exudates and microaerobiosis (Matilla *et al.*, 2011). It will be interesting to investigate whether other DGCs from Sm8530 are involved in the production of this new ML β-glucan and the conditions that might affect their transcription or activity. A good candidate to start this investigation would be SMb20389, a protein displaying GGDEF and GAF domains located right next to the *bgsBA* operon.

The global transcription analysis of the Δ*bgrU* strain has uncovered very interesting evidences for future studies. First, it seems that BgrU potentially regulates three exopolysaccharides: the ML β-glucan, the EPS II, and a putative new one produced by

SMb20460. The repression of the EPS II genes in the Δ*bgrU* strain is consistent with the observed phenotypes regarding mucoidy, aggregation in static conditions, adhesion to polystyrene micro plates, and total EPS II quantification. We hypothesize that, in response to specific environmental signals sensed by BgrU, part of the carbohydrate source is redirected towards the synthesis of the ML β-glucan, which would be more effective in certain niches of the root surface. The fact that *bgsA* (the ML β-glucan synthase) is constitutively expressed thanks to ExpR, but its activity depends on c-di-GMP levels, suggest that BgsA is present in the cell in a latent or basal status, ready to give a quick response whenever is required. We must take into account that the EPS II machinery seems to be only attenuated, since the microarray data showed only 10 down-regulated genes out of the 22 that constitute the *exp* operons. Furthermore, none of the known regulators of EPS II synthesis such as *wgg*R, *muc*R or *phoB* (Bahlawane *et al.*, 2008) were found to be downregulated in the Δ*bgrU* strain, suggesting that the EPS II machinery would be also prepared to start the production again when required. In relation to this hypothesis, new questions arise when we analyze some of the double mutants from the *bgr* operon. Those mutants that produce the ML β-glucan need the presence of the DGC BgrR, otherwise they lose their CF^+ and CR^+ phenotypes, but they also recover the mucoid aspect of the wild type, suggesting an increment of the EPS II synthesis. How is this mucoid reversion achieved? Would it be related to changes in c-di-GMP due to the absence of BgrR, or could there be additional regulatory elements that we ignore? As mentioned before, there is a remarkably number of EPS that have been proved to be regulated by c-di-GMP, but we have not found any reference in relation to EPS II. We believe a good starting point to unravel these questions would be to measure and compare the transcription levels of the *exp* genes in different genetic backgrounds such as Sm8530, Δ*bgrR*, Δ*bgrU*, Δ*bgrV,* Δ*bgrR*/Δ*bgrU* and Sm8530 overexpressing *bgrR* in order to detect any effect regarding the presence/absence of BgrR.

SMb20460, whose gene is induced in the Δ*bgrU* strain, is the other protein with a specific glycosyltransferase domain annotated as a cellulose synthase besides BgsA. Interestingly, it is located closer to the *bgr* operon (around 14 Kb downstream). Together with the up-stream genes *uxs1*and *uxe*, they are predicted to constitute an operon (Schluter *et al.*, 2013). *Uxs1* and *uxe* code for NAD(P) binding proteins involved in carbohydrate modifications. The first one is a dTDP-glucose 4,6-dehydratase and the second protein is an UDP-glucose 4-epimerase (Marchler-Bauer *et al.*, 2015). Additionally, the next three genes downstream *SMb20460* are predicted to form another operon and are also induced in

the Δ*bgrU* strain. Two of them are hypothetical proteins (*SMb20461* and *SMb20463*) and the other one codes for a putative glycosyl hydrolase (Marchler-Bauer *et al.*, 2015). Thus, it seems this set of genes might be implicated in the synthesis of a polysaccharide. However, we only have a very preliminary result showing morphological differences between the fibrils formed by the Δ*bgrU* strain and the ones produced by a *bgrU*/*SMb20460* double mutant. We ignore whether SMb20460 plays an unidentified role in the ML β-glucan synthesis, or it is producing another unknown independent polysaccharide.

In addition to polysaccharides related genes, a mutation in *bgrU* seems to influence the transcription of more than 100 genes, being the majority up-regulated. It is worthy to distinguish that a significant number of the total genes that are differentially expressed are clustered in known/predicted operons, or either are located in the same genetic region, which gives consistency to the analysis. We should highlight the set of up-regulated genes in the pSymA megaplasmid, related to rhizobactin metabolism, denitrification and universal stress response proteins (Usr). Regarding the siderophore rhizobactin 1021, up-regulated genes include those involved in the synthesis (*rhbABCEF* and *SMa2339*), the outer membrane receptor protein (*rhtA)*, and a specific permease (*rhtX*) (Cuiv *et al.*, 2004, Lynch *et al.*, 2001). Besides rhizobactin 1021, there are also chromosome up-regulated genes, *SMc02726* and *hmuSTV,* which code for a hemin-binding regulated outer membrane protein, and ABC-transporter implicated in the haem uptake as a source of iron, respectively (Wexler *et al.*, 2001, Viguier *et al.*, 2005). Siderophores are low-molecular-weight chelators produced commonly by soil microorganisms that show specific affinity for Fe (III) and are synthesized only under iron-limiting conditions (Crosa & Walsh, 2002). It has been proved that during the symbiosis process there is an important iron requirement (Appleby, 1984, Ohara *et al.*, 1988). However, the *rhbABCDEF* operon and *rhtA* are not expressed when the nitrogenase is fully active, suggesting that rhizobactin 1021 might be relevant during free-living or early symbiotic stages of the bacteria (Lynch *et al.*, 2001). Aditionally, Nogales *et al.* (2010), demonstrated that rhizobactin 1021 genes are involved in swarming motility in *S. meliloti* 1021.

In relation to denitrification, the following genes were found to be up-regulated in the Δ*bgrU* strain: almost all genes from the operon encoding a nitrous oxide reductase (*nosZDFYL), nnrU*, necessary for the expression of the *nir* and *nor* genes; *napB* and *nirV* from the *nap* and *nir* operons involved in nitrate-dependent growth (Torres *et al.*, 2011). The activity of denitrification genes in *S. meliloti* has been recently demonstrated.

Interestingly, *S. melilti* was unable to grow using nitrate respiration when cells were incubated in anoxic conditions from the beginning, but not if they were initially incubated with 2% O_2 (Torres *et al.*, 2011). The authors suggest that the cells need a transition stage for a minimum time in a micro aerobic environment to generate all the possible ATP, via aerobic respiration, for the synthesis of the denitrification proteins (Torres *et al.*, 2014). Consistently, genes *fixP1* and *fixQ1* (from the first copy of the fixNOQP operon) are also induced in the Δ*bgrU* strain. These genes code for the high-affinity cytochrome cbb_3 terminal oxidase involved in ATP generation under micro aerobic conditions (Torres *et al.*, 2011). Other cytochrome associated genes are also up-regulated in the Δ*bgrU* strain (*SMb21489, SMa1170*). It will be interesting to perform analysis with the Δ*bgrU* strain in low oxygen conditions to verify whether it presents any advantage using nitrate as final electron acceptors.

In relation to the stress response, we found six up-regulated genes coding for putative universal stress response family (Usr): *SMa1095, SMa1100, SMa1147, SMa1149, SMa1158,* and *SMa1231*. The latter four seem to be target of FixJ, a major regulator of a wide range of processes like denitrification, amino acid/polyamine metabolism and transport, or nitrogenase synthesis (Bobik *et al.*, 2006). However, the concrete stress response in which these small proteins are involved have not been elucidated.

Two putative transcription regulators are induced in the Δ*bgrU* mutant, *SMb20162* and *SMb20344*. The former belongs to the LuxR-type family and it has been reported to be induced with alfalfa root exudates and apigenin (a *nod* gene inducer) at the early stages of nodulation (Zhang & Cheng, 2006). The latter is an AraC-type transcription regulator that has been described to be induced under carbon starvation an osmotic stress (Krol & Becker, 2011).

As a summary, we did not find any evidences suggesting a role of the ML β-glucan in symbiosis, although it seems clear its importance for the root attachment. Thus, it is reasonable to think that the set of genes induced by the *bgrU* mutation regarding denitrification, micro aerobic respiration, rhizobactin synthesis, etc. are expressed during the free-living stage of the cell. Additionally, the chemical/physical properties of the ML β-glucan and the floccules it forms have yet to be studied in detail. Maybe the fibrils formed by this EPS provide the cells a very strong protection but, as a consequence, they reduce the oxygen, iron or carbon availability inside the floccules, so the cells must increase their resources to survive. In this case the proper environmental conditions inside the fibrils

would trigger the induction of the genes above mentioned. On the other hand, we have seen that the ML β-glucan seems to be relevant along the whole process of attachment to the root, in contrast to EPS II, which is determinant at the very beginning. The ML β-glucan might be useful where bacteria reach more hidden micro niches along the root where the oxygen availability or other nutrients could be scarce. We should highlight that Δ*bgrU* and Δ*bgV* mutants seems to growth better than the wild type in calcium depleted medium. Considering the whole set of genes that are induced in the Δ*bgrU* strain, it seems that BgrU could respond to stimulus related to micro-oxic conditions and iron depletion. Nevertheless, the expression of the genes differentially regulated in the Δ*bgrU* strain should be verified first by Q-RT-PCR before continuing with further experiments. Additionally, a closer look into the environmental signals affecting the phosphatase activity of BgrU will help in a better understanding of the role that this transduction signal system plays in the living style of the *Sinorhizobium meliloti*.

CONCLUSIONES

1. La proteína SMb20447 es una diguanilato ciclasa functional cuya actividad induce la producción del exopolisacárido β-glucano de enlace mixto en *Sinorhizobium meliloti* a través de la glucano sintasa BgsA.

2. *SMb20447* forma parte de un operón compuesto por otros cinco genes, desde *SMb20448* hasta *SMb20452*. Las proteínas correspondientes que codifican constituyen un sistema que controla en último término la actividad diguanilato ciclasa de *SMb20447*. Este sistema se asemeja a los conocidos como "partner switching" formados por proteínas fosfatasas, serine quinasas y STAS. Por ello, hemos renombrado el operón de estudio como *bgrRSTUWV* (β-glucan regulation).

3. La transcripción de los genes *bgsBA* depende del sistema de quorum Sin/ExpR. Sin embargo, el operón *bgr* no está regulado transcripcionalmente por ExpR, por lo que constituye un sistema independiente que modula a nivel post-transcripcional la síntesis del β-glucano de enlace mixto.

4. El estado de fosforilación de BgrV (SMb20452), determinado por el balance entre las actividades de la fosfatasa BgrU (SMb20450) y la quinasa BgrW (SMb20451), es el factor clave del sistema de "partner switching". BgrV no fosforilado inhibe o reduce drásticamente la actividad diguanilato ciclasa de BgrR (SMb20447), mientras que la forma fosforilada permite dicha actividad. Ésta es la primera vez que se describe un sistema de "partner switching" de estas características que regula la actividad de una diguanilato ciclasa.

5. Hemos confirmado resultados previos en relación a la función biológica del β-glucano de enlace mixto. Este nuevo exopolisacárido no substituye al EPS I ni al EPS II durante el proceso simbiótico, pero juega un importante papel en la adhesión a las raíces de alfalfa. Sin embargo, la producción constitutiva del β-glucano de enlace mixto dada por una mutación en *bgrU* (*SMb20450*) no proporciona ninguna ventaja en relación a este tipo de adhesión.

CONCLUSIONS

1. The protein SMb20447 is a functional diguanylate cyclase, whose activity triggers the production of the mixed linkage β-glucan exopolysaccharide of *Sinorhizobium meliloti* through the glucan synthase BgsA.

2. *SMb20447* forms part of an operon comprising five other genes, from *SMb20448* to *SMb20452*. Their corresponding proteins constitute a system that controls in last term the diguanylate cyclase activity of SMb20447. Given the similarity of this system with well-known partner switching complexes containing phosphatase, serine kinase and STAS proteins, we have renamed the operon as *bgrRSTUWV* (β-glucan regulation).

3. The transcription of the *bgsBA* genes is dependent on the Sin/ExpR quorum sensing system. However, the *bgr* operon is not transcriptionally regulated by ExpR, conforming and independent system that regulates at the posttranscriptional level the synthesis of the mixed linkage β-glucan.

4. The phosphorylation status of BgrV (SMb20452), determined by the activity balance between the phosphatase BgrU (SMb20450) and the kinase BgrW (SMb20451), is the key factor of the partner switching system. Non-phosphorylated BgrV inhibits or drastically reduces the diguanylate cyclase activity of BgrR (SMb20447), whereas the phosphorylated form allows such activity. To our knowledge, this is the first report of a partner switching mechanism of these characteristics that regulates the activity of a diguanylate cyclase.

5. We confirmed previous results regarding the biological role of the mixed linkage β-glucan. This novel exopolysaccharide does not substitute EPS I nor EPS II in the symbiotic process, but it has proved to be important for the attachment to alfalfa root surfaces. However, the constitutive production of the mixed linkage β-glucan provided by a mutation in *bgrU* (SMb20450) does not present any advantage in the attachment to the root.

SUPPLEMENTARY MATERIAL

Supplementary Table 1. List of *S. meliloti* 2011 mini-Tn*5* mutants obtained from Anke Becker library.

Library mutant ID	Abbreviation	Gene ID	Replicon	Size		mTn*5* location		mTn*5* orientation
				Nucleotides	**Amino acids**	**Nucleotide**	**Amino acid**	
2011mTn5STM.2.13.A06	A1	*SMa0137*	pSymA	2198	733	96	32	Antisense
2011mTn5STM.2.08.C11	A2a	*SMa1548*	pSymA	3212	1071	2073	691	Sense
2011mTn5STM.3.10.D12	A2b					782	260	Sense
2011mTn5STM.1.04.E01	B1	*SMb20389*	pSymB	1022	341	599	199	Antisense
2011mTn5STM.4.02.F08	B2a	*SMb20447*	pSymB	1691	564	1229	409	Antisense
2011mTn5STM.4.04.E09	B2b					140	46	Antisense
2011mTn5STM.2.06.E10	B4a	*SMb20900*	pSymB	1931	644	1869	623	Sense
2011mTn5STM.2.11.H01	B4b					1616	538	Antisense
2011mTn5STM.4.02.E09	C1	*SMc00887*	Chromosome	1343	448	432	144	Sense
2011mTn5STM.2.08.E05	C4	*SMc00033*	Chromosome	1796	599	1736	578	Sense
2011mTn5STM.4.05.D12	C5a	*SMc00038*	Chromosome	2315	772	200	66	Sense
2011mTn5STM.2.10.C08	C5b					682	227	Antisense
2011mTn5STM.4.02.H01	C5c					1479	493	Antisense
2011mTn5STM.4.05.E12	C8a	*SMc03942*	Chromosome	2318	773	2231	743	Antisense
2011mTn5STM.2.12.E10	C8b					291	97	Sense
2011mTn5STM.3.02.E02	C9	*SMc03178*	Chromosome	2645	882	175	58	Sense
2011mTn5STM.4.11.A03	C10	*SMc03141*	Chromosome	1046	349	70	23	Antisense

Supplementary Table 2. Mutants, complementations, and crossed complementations with their corresponding CF phenotypes in TY plates. (++) denotes a more intense fluorescence than (+).

Strain/ Mutant	Presence/Absence of the native gene in pSymB						Gene expressed in *trans*	CF phenotype
	SMb20447	*SMb20448*	*SMb20449*	*SMb20450*	*SMb20451*	*SMb20452*		
Sm8530	✓	✓	✓	✓	✓	✓	-	-
Single mutants	×	✓	✓	✓	✓	✓	-	-
	✓	×	✓	✓	✓	✓	-	-
	✓	✓	×	✓	✓	✓	- *SMb20449*	+ -
	✓	✓	✓	×	✓	✓	- *SMb20450* *SMb20452* *SMb20452* (*S61A*) *SMb20447* *SMb20447* (E215A)	++ - ++ - ++ -
	✓	✓	✓	✓	×	✓*	- *SMb20451* *SMb20452*	+ ++ -
	✓	✓	✓	✓	✓	×	- *SMb20452* *SMb20450*	++ - ++
Overexpressions	✓	✓	✓	✓	✓	✓	*SMb20447* *SMb20447* (E215A)	++ -
	✓	✓	✓	✓	✓	✓	*SMb20448*	+
	✓	✓	✓	✓	✓	✓	*SMb20449*	-
	✓	✓	✓	✓	✓	✓	*SMb20450*	-
	✓	✓	✓	✓	✓	✓	*SMb20451*	++
	×	✓	✓	✓	✓	✓		-
	✓	✓	✓	✓	✓	✓	*SMb20452*	-
Double/Triple mutants	×	✓	×	✓	✓	✓	- *SMb20447*	- ++
	×	✓	✓	×	✓	✓	- *SMb20447*	- ++
	×	✓	✓	✓	✓	×	- *SMb20447*	- ++
	✓	✓	✓	×	✓	×	- *SMb20452*	++ ++
	✓	✓	✓	×	×	×	- *SMb20452*	++ -

(*)Possibility of considering the Δ*SMb20451* mutant as a double mutant *SMb20451*/*SMb20452* due to the putative polar effect of the deletion in *SMb20451*.

Supplementary Table 3. List of up-regulated genes in the Δ*SMb20450* mutant strain.

Gene	Description	M value
SMa0661	Hypothetical protein	1,044325038
SMa0675	Calcium/Proton antiporter	1,35392837
SMa0678	Putrescine/Ornithine antiporter	1,596673675
SMa0680	Amino acid (ornithine, lysine, arginine) decarboxylase	1,088466943
SMa0682	Amino acid (ornithine, lysine, arginine) decarboxylase	1,44804421
SMa0689	Hypothetical protein	1,192019012
SMa0690	Hypothetical protein	1,423739763
SMa0695 (*arcB*)	Ornithine carbamoyltransferase	1,142397226
SMa0763	Pyridoxamine 5'-phosphate oxidase-related FMN-binding protein	1,136918647
SMa0767 (*fixQ2*)	Cbb3-type cytochrome oxidase component	1,348974792
SMa1084	D-xylulose 5-phosphate/D-fructose 6-phosphate phosphoketolase	1,035810696
SMa1086	CBS (Cystathionine Beta-Synthase) and BON (Bacterial OsmY and Nodulation) domain containing protein.	1,127109333
SMa1095	Universal stress protein (Usp)	1,087590777
SMa1100	Universal stress protein (Usp)	1,030793014
SMa1126	Zinc metalloprotease domain containing protein.	1,057893594
SMa1128 (*degP4*)	Trypsin-like serine protease	1,207022721
SMa1131	RNA-metabolism metallo-beta-lactamase	1,006758508
SMa1146	Succinyl-CoA synthetase-like	1,119299119
SMa1147	Universal stress protein (Usp)	1,159981168
SMa1149	Universal stress protein (Usp)	1,469027641
SMa1151	Pyridoxamine 5'-phosphate oxidase-related FMN-binding protein	1,495045921
SMa1153	Bacterial OsmY and nodulation domain containing protein	1,010264818

SMa1154	Hypothetical protein	1,193936932
SMa1155	Cation transporter / Haloacid dehalogenase	1,363406871
SMa1156	Alcohol dehydrogenase	1,524258002
SMa1158	Universal stress protein (Usp)	1,384717459
SMa1170	Cytochrome c, monohaem.	1,050967332
SMa1182 (*nosZ*)	Nitrous oxide reductase	1,071442991
SMa1183 (*nosD*)	Nitrous oxidase accessory protein	1,536351195
SMa1184 (*nosF*)	ABC transporter, ATP-binding protein	1,30686542
SMa1185 (*nosY*)	ABC-type transport system involved in multi-copper enzyme maturation, permease component	1,373197233
SMa1186 (*nosL*)	Nitrous oxide reductase accessory protein, putative copper chaperone	1,042628653
SMa1209 (*fixI1*)	Copper-translocating P-type ATPase	1,049686934
SMa1213 (*fixP1*)	Cytochrome C oxidase, cbb3-type	1,375278593
SMa1214 (*fixQ1*)	Cytochrome C oxidase cbb3-type component	1,388503958
SMa1231	Universal stress protein (Usp)	1,09276057
SMa1233 (*napB*)	Nitrate reductase cytochrome c-type subunit	1,012883781
SMa1247 (*nirV*)	Accessory protein for dissimilatory nitrite reduction	1,087551574
SMa1283 (*nnrU*)	Denitrification regulatory protein	1,205855711
SMa1285	3-octaprenyl-4-hydroxybenzoate carboxy-lyase	1,231214068
SMa1296 (*adhA1*)	Alcohol dehydrogenase	1,341256997
SMa2337 (*rhtX*)	Rhizobactin transporter	1,887710773
SMa2339	Siderophore biosynthesis	1,224486232
SMa2400 (*rhbA*)	Diaminobutyrate--2-oxoglutarate aminotransferase involved in rhizobactin biosynthesis	1,548697523
SMa2402 (*rhbB*)	L-2,4-diaminobutyrate decarboxylase	1,158701083

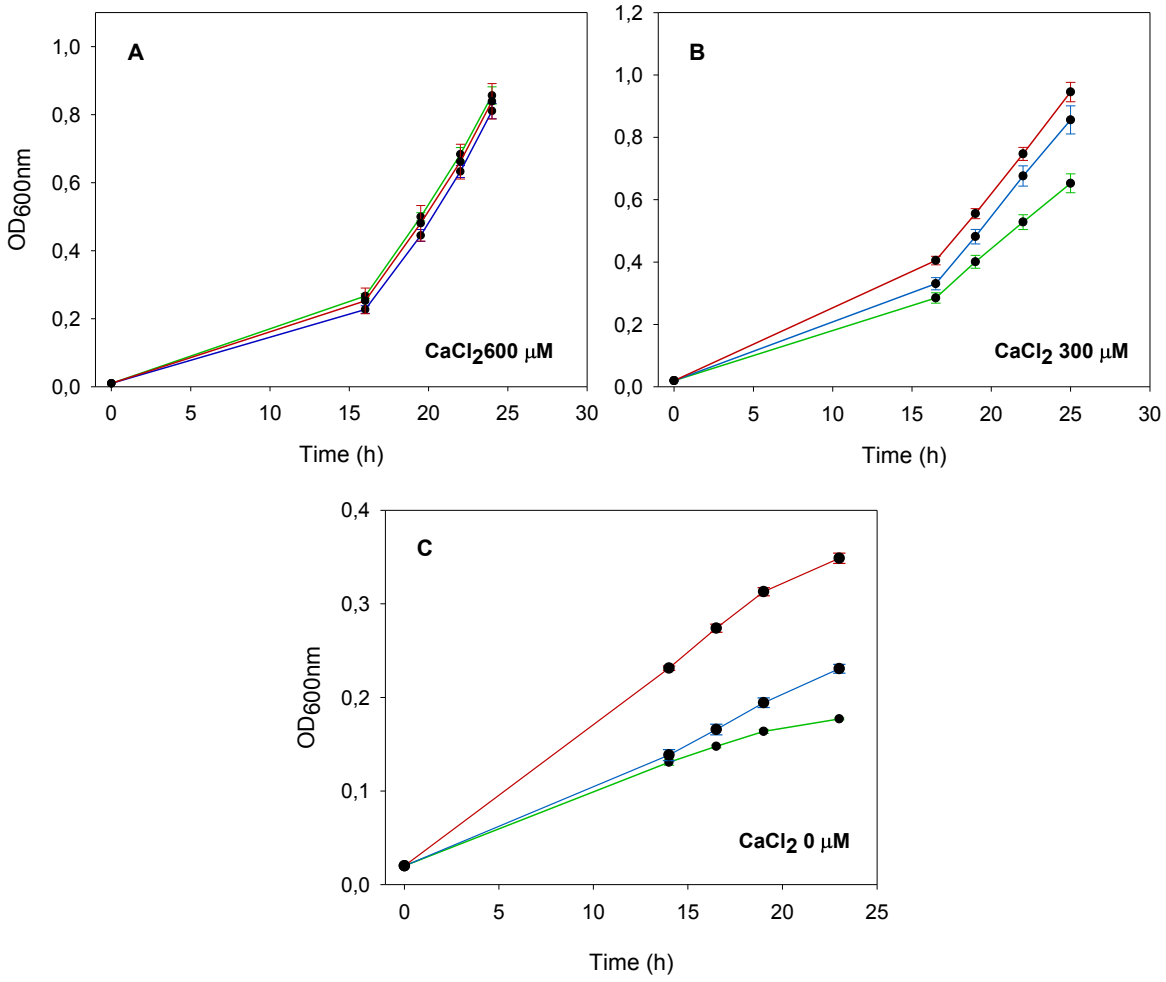
SMa2404 (*rhbC*)	Rhizobactin biosynthesis	1,111386363
SMa2408 (*rhbE*)	Rhizobactin biosynthesis	1,32007843
SMa2410 (*rhbF*)	Rhizobactin biosynthesis	1,264218187
SMa2414 (*rhtA*)	Rhizobactin receptor	1,691518151
SMb20025	Hypothetical protein	1,262360663
SMb20162	LuxR-type transcriptional regulator	1,081675348
SMb20344	AraC-type transcriptional regulator	1,231969226
SMb20460	Glycosyl transferase (putative cellulose synthase)	1,92792948
SMb20461	Hypothetical protein	1,051162501
SMb20462	Glycoside hydrolase/beta-mannosidase	2,547743318
SMb20463	Hypothetical protein	1,435913597
SMb20611 (*dctA*)	C4-dicarboxylate transporter	1,217019045
SMb20842	TonB-dependent receptor	1,101309141
SMb20906	Hypothetical protein	1,116664367
SMb21154	Glyoxalase/bleomycin resistance protein/dioxygenase domain containing protein	1,21168652
SMb21456	Hypothetical protein	1,055576727
SMb21489 (*cyoC*)	Ubiquinol oxidase subunit III	1,014016516
SMc00092 (*cysH*)	Phosphoadenosine phosphosulphate reductase	1,014536225
SMc00506	Hypothetical membrane protein	1,28365374
SMc00591	Hypothetical protein	1,063380159
SMc00653	Putative 2-component receiver domain protein	1,254404813
SMc00784 (*fbpA*)	Fe^{3+} ABC transporter	1,087797071
SMc00854	Hypothetical protein	1,176504317
SMc00987	Hypothetical protein	1,004205506

SMc01488	Hypothetical membrane protein	1,948964618
SMc01510 (*hmuV*)	Hemin transport system, ABC transporter	1,103124681
SMc01512 (*hmuT*)	Hemin binding periplasmic transmembrane protein	1,151340032
SMc01513 (*hmuS*)	Hemin degrading protein	1,886649458
SMc01514	Hypothetical protein	1,135064236
SMc01589	Hypothetical protein	1,083332752
SMc01600	Putative N-formylglutamate amidohydrolase	1,040263656
SMc01626	Putative polyol ABC transporter, permease component	1,010289642
SMc01780 (*hppA*)	H^+ translocating membrane pyrophosphatase	1,043649938
SMc02177	Hypothetical protein	1,1841735
SMc02178	Hypothetical protein	1,276601279
SMc02227(*fadB*)	Putative fatty oxidation complex alpha subunit	1,031959032
SMc02229	Acyl-CoA dehydrogenase	1,239681425
SMc02254 (*qxtB*)	Cytochrome bd ubiquinol oxidase, subunit II	1,024792777
SMc02689	Aldehyde dehydrogenase	1,09417106
SMc02726 (*shmR*)	TonB-dependent haemoglobin/transferrin/lactoferrin receptor	1,394263178
SMc03787	Hypothetical protein	1,117453902
SMc04173	Putative dihydrofolate reductase	1,044125961
SMc04190	Hypothetical protein	1,429913905

Supplementary Table 4. List of genes down-regulated in the Δ*SMb20450* mutant strain.

Gene	Description	M value
SMa0585 (*nrtA*)	Nitrate transporter	-1,13183844
SMa0590	Hypothetical protein	-1,42837326
SMa0591	Transposase fragment	-1,18982553
SMa0983	Hypothetical protein	-1,39586881
SMa1081	Hypothetical protein	-1,78319568
SMa1550	LuxR-type transcriptional regulator	-1,05969821
SMa1874	DNA breaking/re-joining enzyme, catalytic core	-1,22595451
SMa2267	Glyoxalase/bleomycin resistance/dioxygenase domain containing protein	-1,21573776
SMb20037 (*aroE2*)	Shikimate dehydrogenase	-2,12155588
SMb20089	Hypothetical protein	-2,26004418
SMb20174	Cytochrome c	-1,06025765
SMb20436	MFS transporter (putative nitrate transporter)	-1,6421418
SMb20745 (*glnII*)	Glutamine synthetase	-1,10721446
SMb20822 (*rkpT2*)	Putative cell surface polysaccharide export, ABC-2 transporter permease	-1,81025003
SMb20974	Oxidoreductase	-1,13918101
SMb20984 (*nirB*)	Nitrite reductase large subunit	-1,90951499
SMb21080	LuxR-type transcriptional regulator	-1,05616567
SMb21290	NADPH quinone oxidoreductase	-1,05704549
SMb21307 (*wgeH*)	Hypothetical protein	-1,38161954
SMb21312 (*wgeC*)	Type 11 methyltransferase	-2,10490416
SMb21313 (*wgeB*)	Putative bifunctional glycosyltransferase	-1,85825683
SMb21314 (*wgeA*)	Hypothetical calcium binding protein	-1,47258395
SMb21315 (*wgdB*)	Putative secretion membrane fusion protein	-1,11129195

SMb21319 (*wgaA*)	Putative membrane-anchored protein	-1,6646575
SMb21320 (*wgaB*)	Glycosyl transferase, family 1	-1,19538752
SMb21321 (*wgaD*)	Polysaccharide pyruvyl transferase (putative membrane-anchored protein)	-1,20744056
SMb21323 (*wgaF*)	Glycoside hydrolase	-1,46271942
SMb21327 (*wgaJ*)	Putative dTDP-4-dehydrorhamnose reductase	-1,52220968
SMb21604	Periplasmic-binding component of ABC transport system specific for sn-glycerol-3-phosphate	-1,0025887
SMc00428	Putative transmembrane protein	-1,02629568
SMc00521	NUDIX type hydrolase	-1,06795083
SMc01313 (*rpsG*)	30S ribosomal protein S7	-1,11246902
SMc02084 (*exbD*)	Putative biopolymer transport protein	-1,00086442
SMc03242 (*typA*)	Ribosome-binding GTPase	-1,06719318
SMc03806 (*glnK*)	Nitrogen regulatory protein PII	-1,96384285
SMc03807(*amtB*)	Ammonium transporter	-1,70593045
SMc04117	TadE-like protein (hypothetical pilus assembly)	-1,01320203
SMc04170	Two-component system response regulator	-1,41791065
SMc04171	Putative hemolysin-type calcium-binding protein	-2,57124395



Supplementary Figure 1. Growth of strains Sm8530 (green line), Δ*SMb20450* (blue line) and Δ*SMb20452* (red line) in TY medium supplemented with different $CaCl_2$ concentrations.

REFERENCES

Adesemoye, A. O., H. A. Torbert & J. W. Kloepper, (2009) Plant growth-promoting rhizobacteria allow reduced application rates of chemical fertilizers. *Microbial ecology* **58**: 921-929.

Anantharaman, V. & L. Aravind, (2000) Cache - a signaling domain common to animal Ca(2+)-channel subunits and a class of prokaryotic chemotaxis receptors. *Trends in biochemical sciences* **25**: 535-537.

Appleby, C. A., (1984) LEGHEMOGLOBIN AND RHIZOBIUM RESPIRATION. *Annu. Rev. Plant Physiol. Plant Molec. Biol.* **35**: 443-478.

Aravind, L. & C. P. Ponting, (1999) The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS microbiology letters* **176**: 111-116.

Ausmees, N., R. Mayer, H. Weinhouse, G. Volman, D. Amikam, M. Benziman & M. Lindberg, (2001) Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. *FEMS microbiology letters* **204**: 163-167.

Bahlawane, C., B. Baumgarth, J. Serrania, S. Ruberg & A. Becker, (2008) Fine-tuning of galactoglucan biosynthesis in Sinorhizobium meliloti by differential WggR (ExpG)-, PhoB-, and MucR-dependent regulation of two promoters. *Journal of bacteriology* **190**: 3456-3466.

Barnett, M. J., R. F. Fisher, T. Jones, C. Komp, A. P. Abola, F. Barloy-Hubler, L. Bowser, D. Capela, F. Galibert, J. Gouzy, M. Gurjal, A. Hong, L. Huizar, R. W. Hyman, D. Kahn, M. L. Kahn, S. Kalman, D. H. Keating, C. Palm, M. C. Peck, R. Surzycki, D. H. Wells, K. C. Yeh, R. W. Davis, N. A. Federspiel & S. R. Long, (2001) Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti pSymA megaplasmid. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 9883-9888.

Bassler, B. L., (1999) How bacteria talk to each other: regulation of gene expression by quorum sensing. *Current opinion in microbiology* **2**: 582-587.

Battisti, L., J. C. Lara & J. A. Leigh, (1992) Specific oligosaccharide form of the Rhizobium meliloti exopolysaccharide promotes nodule invasion in alfalfa. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 5625-5629.

Bauer, W. D. & U. Mathesius, (2004) Plant responses to bacterial quorum sensing signals. *Current opinion in plant biology* **7**: 429-433.

Becker, A., (2015) Challenges and perspectives in combinatorial assembly of novel exopolysaccharide biosynthesis pathways. *Frontiers in microbiology* **6**: 687.

Becker, A., S. Ruberg, B. Baumgarth, P. A. Bertram-Drogatz, I. Quester & A. Puhler, (2002) Regulation of succinoglycan and galactoglucan biosynthesis in Sinorhizobium meliloti. *Journal of molecular microbiology and biotechnology* **4**: 187-190.

Becker, A., S. Ruberg, H. Kuster, A. A. Roxlau, M. Keller, T. Ivashina, H. P. Cheng, G. C. Walker & A. Puhler, (1997) The 32-kilobase exp gene cluster of Rhizobium meliloti

directing the biosynthesis of galactoglucan: genetic organization and properties of the encoded gene products. *Journal of bacteriology* **179**: 1375-1384.

Becker, A., M. Schmidt, W. Jager & A. Puhler, (1995) New gentamicin-resistance and lacZ promoter-probe cassettes suitable for insertion mutagenesis and generation of transcriptional fusions. *Gene* **162**: 37-39.

Becker A., R. S., (2003) S. meliloti Sm6kOligo Microarray Manual, V 1.2, University of Bielefeld

Bertram-Drogatz, P. A., I. Quester, A. Becker & A. Puhler, (1998) The Sinorhizobium meliloti MucR protein, which is essential for the production of high-molecular-weight succinoglycan exopolysaccharide, binds to short DNA regions upstream of exoH and exoY. *Molecular & general genetics : MGG* **257**: 433-441.

Bobik, C., E. Meilhoc & J. Batut, (2006) FixJ: a major regulator of the oxygen limitation response and late symbiotic functions of Sinorhizobium meliloti. *Journal of bacteriology* **188**: 4890-4902.

Bobrov, A. G., O. Kirillina, D. A. Ryjenkov, C. M. Waters, P. A. Price, J. D. Fetherston, D. Mack, W. E. Goldman, M. Gomelsky & R. D. Perry, (2011) Systematic analysis of cyclic di-GMP signalling enzymes and their role in biofilm formation and virulence in Yersinia pestis. *Molecular microbiology* **79**: 533-551.

Bogino, P. C., L. Oliva Mde, F. G. Sorroche & W. Giordano, (2013) The role of bacterial biofilms and surface components in plant-bacterial associations. *International journal of molecular sciences* **14**: 15838-15859.

Bradford, M. M., (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* **72**: 248-254.

Breedveld, M. W., H. C. Cremers, M. Batley, M. A. Posthumus, L. P. Zevenhuizen, C. A. Wijffelman & A. J. Zehnder, (1993) Polysaccharide synthesis in relation to nodulation behavior of Rhizobium leguminosarum. *Journal of bacteriology* **175**: 750-757.

Brown, T., (2001) Southern blotting. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **Chapter 2**: Unit2 9A.

Caly, D. L., D. Bellini, M. A. Walsh, J. M. Dow & R. P. Ryan, (2015) Targeting cyclic di-GMP signalling: a strategy to control biofilm formation? *Current pharmaceutical design* **21**: 12-24.

Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon & M. Robles, (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674-3676.

Costerton, J. W., Z. Lewandowski, D. E. Caldwell, D. R. Korber & H. M. Lappin-Scott, (1995) Microbial biofilms. *Annual review of microbiology* **49**: 711-745.

Crosa, J. H. & C. T. Walsh, (2002) Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiology and molecular biology reviews : MMBR* **66**: 223-249.

Cuiv, P. O., P. Clarke, D. Lynch & M. O'Connell, (2004) Identification of rhtX and fptX, novel genes encoding proteins that show homology and function in the utilization of the siderophores rhizobactin 1021 by Sinorhizobium meliloti and pyochelin by Pseudomonas aeruginosa, respectively. *Journal of bacteriology* **186**: 2996-3005.

Chan, C., R. Paul, D. Samoray, N. C. Amiot, B. Giese, U. Jenal & T. Schirmer, (2004) Structural basis of activity and allosteric control of diguanylate cyclase. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 17084-17089.

Diebold, R. & K. D. Noel, (1989) Rhizobium leguminosarum exopolysaccharide mutants: biochemical and genetic analyses and symbiotic behavior on three hosts. *Journal of bacteriology* **171**: 4821-4830.

Dilanji, G. E., M. Teplitski & S. J. Hagen, (2014) Entropy-driven motility of Sinorhizobium meliloti on a semi-solid surface. *Proceedings. Biological sciences / The Royal Society* **281**: 20132575.

Djordjevic, S. P., H. Chen, M. Batley, J. W. Redmond & B. G. Rolfe, (1987) Nitrogen fixation ability of exopolysaccharide synthesis mutants of Rhizobium sp. strain NGR234 and Rhizobium trifolii is restored by the addition of homologous exopolysaccharides. *Journal of bacteriology* **169**: 53-60.

Dombrecht, B., J. Vanderleyden & J. Michiels, (2001) Stable RK2-derived cloning vectors for the analysis of gene expression and gene function in gram-negative bacteria. *Molecular plant-microbe interactions : MPMI* **14**: 426-430.

Dondrup, M., A. Goesmann, D. Bartels, J. Kalinowski, L. Krause, B. Linke, O. Rupp, A. Sczyrba, A. Puhler & F. Meyer, (2003) EMMA: a platform for consistent storage and efficient analysis of microarray data. *Journal of biotechnology* **106**: 135-146.

Dubois, M., K. Gilles, J. K. Hamilton, P. A. Rebers & F. Smith, (1951) A colorimetric method for the determination of sugars. *Nature* **168**: 167.

Dutta, R. & M. Inouye, (2000) GHKL, an emergent ATPase/kinase superfamily. *Trends in biochemical sciences* **25**: 24-28.

Eberhard, A., A. L. Burlingame, C. Eberhard, G. L. Kenyon, K. H. Nealson & N. J. Oppenheimer, (1981) Structural identification of autoinducer of Photobacterium fischeri luciferase. *Biochemistry* **20**: 2444-2449.

Engebrecht, J., R. Brent & M. A. Kaderbhai, (2001) Minipreps of plasmid DNA. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **Chapter 1**: Unit1 6.

Fang, X., I. Ahmad, A. Blanka, M. Schottkowski, A. Cimdins, M. Y. Galperin, U. Romling & M. Gomelsky, (2014) GIL, a new c-di-GMP-binding protein domain involved in regulation of cellulose synthesis in enterobacteria. *Molecular microbiology* **93**: 439-452.

Ferreira, R. B., L. C. Antunes, E. P. Greenberg & L. L. McCarter, (2008) Vibrio parahaemolyticus ScrC modulates cyclic dimeric GMP regulation of gene expression relevant to growth on surfaces. *Journal of bacteriology* **190**: 851-860.

Finan, T. M., E. Hartweig, K. LeMieux, K. Bergman, G. C. Walker & E. R. Signer, (1984) General transduction in Rhizobium meliloti. *Journal of bacteriology* **159**: 120-124.

Finan, T. M., B. Kunkel, G. F. De Vos & E. R. Signer, (1986) Second symbiotic megaplasmid in Rhizobium meliloti carrying exopolysaccharide and thiamine synthesis genes. *Journal of bacteriology* **167**: 66-72.

Finan, T. M., S. Weidner, K. Wong, J. Buhrmester, P. Chain, F. J. Vorholter, I. Hernandez-Lucas, A. Becker, A. Cowie, J. Gouzy, B. Golding & A. Puhler, (2001) The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 9889-9894.

Finn, R. D., A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. Sonnhammer, J. Tate & M. Punta, (2014) Pfam: the protein families database. *Nucleic acids research* **42**: D222-230.

Flemming, H. C. & J. Wingender, (2010) The biofilm matrix. *Nature reviews. Microbiology* **8**: 623-633.

Fouhy, Y., J. F. Lucey, R. P. Ryan & J. M. Dow, (2006) Cell-cell signaling, cyclic di-GMP turnover and regulation of virulence in Xanthomonas campestris. *Research in microbiology* **157**: 899-904.

Francez-Charlot, A., A. Kaczmarczyk, H. M. Fischer & J. A. Vorholt, (2015) The general stress response in Alphaproteobacteria. *Trends in microbiology* **23**: 164-171.

Fraysse, N., F. Couderc & V. Poinsot, (2003) Surface polysaccharide involvement in establishing the rhizobium-legume symbiosis. *European journal of biochemistry / FEBS* **270**: 1365-1380.

Freitas, F., V. D. Alves & M. A. Reis, (2011) Advances in bacterial exopolysaccharides: from production to biotechnological applications. *Trends in biotechnology* **29**: 388-398.

Fujishige, N. A., N. N. Kapadia, P. L. De Hoff & A. M. Hirsch, (2006) Investigations of Rhizobium biofilm formation. *FEMS microbiology ecology* **56**: 195-206.

Fuqua, W. C., S. C. Winans & E. P. Greenberg, (1994) Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *Journal of bacteriology* **176**: 269-275.

Gage, D. J., (2004) Infection and Invasion of Roots by Symbiotic, Nitrogen-Fixing Rhizobia during Nodulation of Temperate Legumes. *Microbiology and Molecular Biology Reviews* **68**: 280-300.

Galibert, F., T. M. Finan, S. R. Long, A. Puhler, P. Abola, F. Ampe, F. Barloy-Hubler, M. J. Barnett, A. Becker, P. Boistard, G. Bothe, M. Boutry, L. Bowser, J. Buhrmester, E. Cadieu, D. Capela, P. Chain, A. Cowie, R. W. Davis, S. Dreano, N. A. Federspiel, R. F. Fisher, S. Gloux, T. Godrie, A. Goffeau, B. Golding, J. Gouzy, M. Gurjal, I. Hernandez-Lucas, A. Hong, L. Huizar, R. W. Hyman, T. Jones, D. Kahn, M. L. Kahn, S. Kalman, D. H. Keating, E. Kiss, C. Komp, V. Lelaure, D. Masuy, C. Palm, M. C. Peck, T. M. Pohl, D. Portetelle, B. Purnelle, U. Ramsperger, R. Surzycki, P. Thebault, M. Vandenbol, F. J. Vorholter, S. Weidner, D. H. Wells, K.

Wong, K. C. Yeh & J. Batut, (2001) The composite genome of the legume symbiont Sinorhizobium meliloti. *Science (New York, N.Y.)* **293**: 668-672.

Gallagher, S., S. E. Winston, S. A. Fuller & J. G. Hurrell, (2008) Immunoblotting and immunodetection. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **Chapter 10**: Unit 10 18.

Gallagher, S. R., (2012) One-dimensional SDS gel electrophoresis of proteins. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **Chapter 10**: Unit 10 12A.

Gao, S., S. B. Romdhane, S. Beullens, V. Kaever, I. Lambrichts, M. Fauvart & J. Michiels, (2014) Genomic analysis of cyclic-di-GMP-related genes in rhizobial type strains and functional analysis in Rhizobium etli. *Applied microbiology and biotechnology* **98**: 4589-4602.

Gibson, K. E., H. Kobayashi & G. C. Walker, (2008) Molecular determinants of a symbiotic chronic infection. *Annual review of genetics* **42**: 413-441.

Giddens, S. R., R. W. Jackson, C. D. Moon, M. A. Jacobs, X. X. Zhang, S. M. Gehrig & P. B. Rainey, (2007) Mutational activation of niche-specific genes provides insight into regulatory networks and bacterial function in a complex environment. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 18247-18252.

Glazebrook, J. & G. C. Walker, (1989) A novel exopolysaccharide can function in place of the calcofluor-binding exopolysaccharide in nodulation of alfalfa by Rhizobium meliloti. *Cell* **56**: 661-672.

Glazebrook, J. & G. C. Walker, (1991) Genetic techniques in Rhizobium meliloti. *Methods in enzymology* **204**: 398-418.

Glenn, S. A., N. Gurich, M. A. Feeney & J. E. Gonzalez, (2007) The ExpR/Sin quorum-sensing system controls succinoglycan production in Sinorhizobium meliloti. *Journal of bacteriology* **189**: 7077-7088.

Gonzalez, J. E. & M. M. Marketon, (2003) Quorum sensing in nitrogen-fixing rhizobia. *Microbiology and molecular biology reviews : MMBR* **67**: 574-592.

Gonzalez, J. E., B. L. Reuhs & G. C. Walker, (1996a) Low molecular weight EPS II of Rhizobium meliloti allows nodule invasion in Medicago sativa. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 8636-8641.

Gonzalez, J. E., G. M. York & G. C. Walker, (1996b) Rhizobium meliloti exopolysaccharides: synthesis and symbiotic function. *Gene* **179**: 141-146.

Gray, J. X., H. J. Zhan, S. B. Levery, L. Battisti, B. G. Rolfe & J. A. Leigh, (1991) Heterologous exopolysaccharide production in Rhizobium sp. strain NGR234 and consequences for nodule development. *Journal of bacteriology* **173**: 3066-3077.

Gupta, K., P. Kumar & D. Chatterji, (2010) Identification, activity and disulfide connectivity of C-di-GMP regulating proteins in Mycobacterium tuberculosis. *PloS one* **5**: e15072.

Hazelbauer, G. L. & W. C. Lai, (2010) Bacterial chemoreceptors: providing enhanced features to two-component signaling. *Current opinion in microbiology* **13**: 124-132.

Hecker, M., J. Pane-Farre & U. Volker, (2007) SigB-dependent general stress response in Bacillus subtilis and related gram-positive bacteria. *Annual review of microbiology* **61**: 215-236.

Hengge, R., (2009) Principles of c-di-GMP signalling in bacteria. *Nature reviews. Microbiology* **7**: 263-273.

Hickman, J. W. & C. S. Harwood, (2008) Identification of FleQ from Pseudomonas aeruginosa as a c-di-GMP-responsive transcription factor. *Molecular microbiology* **69**: 376-389.

Hirsch, A. M., M. R. Lum & J. A. Downie, (2001) What makes the rhizobia-legume symbiosis so special? *Plant physiology* **127**: 1484-1492.

Hoang, H. H., A. Becker & J. E. Gonzalez, (2004) The LuxR homolog ExpR, in combination with the Sin quorum sensing system, plays a central role in Sinorhizobium meliloti gene expression. *Journal of bacteriology* **186**: 5460-5472.

Hoang, H. H., N. Gurich & J. E. Gonzalez, (2008) Regulation of motility by the ExpR/Sin quorum-sensing system in Sinorhizobium meliloti. *Journal of bacteriology* **190**: 861-871.

Hogan, D. A., A. Vik & R. Kolter, (2004) A Pseudomonas aeruginosa quorum-sensing molecule influences Candida albicans morphology. *Molecular microbiology* **54**: 1212-1223.

Hua, L., P. S. Hefty, Y. J. Lee, Y. M. Lee, R. S. Stephens & C. W. Price, (2006) Core of the partner switching signalling mechanism is conserved in the obligate intracellular pathogen Chlamydia trachomatis. *Molecular microbiology* **59**: 623-636.

Hunter, J. L., G. B. Severin, B. J. Koestler & C. M. Waters, (2014) The Vibrio cholerae diguanylate cyclase VCA0965 has an AGDEF active site and synthesizes cyclic di-GMP. *BMC microbiology* **14**: 22.

Janczarek, M., (2011) Environmental signals and regulatory pathways that influence exopolysaccharide production in rhizobia. *International journal of molecular sciences* **12**: 7898-7933.

Jones, K. M., H. Kobayashi, B. W. Davies, M. E. Taga & G. C. Walker, (2007) How rhizobial symbionts invade plants: the Sinorhizobium-Medicago model. *Nature reviews. Microbiology* **5**: 619-633.

Kaczmarczyk, A., J. A. Vorholt & A. Francez-Charlot, (2013) Cumate-inducible gene expression system for sphingomonads and other Alphaproteobacteria. *Applied and environmental microbiology* **79**: 6795-6802.

Kanehisa, M., Y. Sato, M. Kawashima, M. Furumichi & M. Tanabe, (2015) KEGG as a reference resource for gene and protein annotation. *Nucleic acids research*.

Karr, D. B., R. T. Liang, B. L. Reuhs & D. W. Emerich, (2000) Altered exopolysaccharides of Bradyrhizobium japonicum mutants correlate with impaired soybean lectin binding, but not with effective nodule formation. *Planta* **211**: 218-226.

Keller, M., A. Roxlau, W. M. Weng, M. Schmidt, J. Quandt, K. Niehaus, D. Jording, W. Arnold & A. Puhler, (1995) Molecular analysis of the Rhizobium meliloti mucR gene regulating the biosynthesis of the exopolysaccharides succinoglycan and galactoglucan. *Molecular plant-microbe interactions : MPMI* **8**: 267-277.

Kim, C. H., R. E. Tully & D. L. Keister, (1989) Exopolysaccharide-Deficient Mutants of Rhizobium fredii HH303 Which Are Symbiotically Effective. *Applied and environmental microbiology* **55**: 1852-1854.

Kinoshita, E., E. Kinoshita-Kikuta & T. Koike, (2015) Advances in Phos-tag-based methodologies for separation and detection of the phosphoproteome. *Biochimica et biophysica acta* **1854**: 601-608.

Kovach, M. E., P. H. Elzer, D. S. Hill, G. T. Robertson, M. A. Farris, R. M. Roop, 2nd & K. M. Peterson, (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**: 175-176.

Kozak, N. A., S. Mattoo, A. K. Foreman-Wykert, J. P. Whitelegge & J. F. Miller, (2005) Interactions between partner switcher orthologs BtrW and BtrV regulate type III secretion in Bordetella. *Journal of bacteriology* **187**: 5665-5676.

Krasteva, P. V., K. M. Giglio & H. Sondermann, (2012) Sensing the messenger: the diverse ways that bacteria signal through c-di-GMP. *Protein science : a publication of the Protein Society* **21**: 929-948.

Krogh, A., B. Larsson, G. von Heijne & E. L. Sonnhammer, (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* **305**: 567-580.

Krol, E. & A. Becker, (2011) ppGpp in Sinorhizobium meliloti: biosynthesis in response to sudden nutritional downshifts and modulation of the transcriptome. *Molecular microbiology* **81**: 1233-1254.

Kumar, M. & D. Chatterji, (2008) Cyclic di-GMP: a second messenger required for long-term survival, but not for biofilm formation, in Mycobacterium smegmatis. *Microbiology* **154**: 2942-2955.

Laus, M. C., T. J. Logman, G. E. Lamers, A. A. Van Brussel, R. W. Carlson & J. W. Kijne, (2006) A novel polar surface polysaccharide from Rhizobium leguminosarum binds host plant lectin. *Molecular microbiology* **59**: 1704-1713.

Laus, M. C., A. A. van Brussel & J. W. Kijne, (2005) Role of cellulose fibrils and exopolysaccharides of Rhizobium leguminosarum in attachment to and infection of Vicia sativa root hairs. *Molecular plant-microbe interactions : MPMI* **18**: 533-538.

Lee, V. T., J. M. Matewish, J. L. Kessler, M. Hyodo, Y. Hayakawa & S. Lory, (2007) A cyclic-di-GMP receptor required for bacterial exopolysaccharide production. *Molecular microbiology* **65**: 1474-1484.

Lefebre, M. D. & M. A. Valvano, (2002) Construction and evaluation of plasmid vectors optimized for constitutive and regulated gene expression in Burkholderia cepacia complex isolates. *Applied and environmental microbiology* **68**: 5956-5964.

Lehman, A. P. & S. R. Long, (2013) Exopolysaccharides from Sinorhizobium meliloti can protect against H2O2-dependent damage. *Journal of bacteriology* **195**: 5362-5369.

Leigh, J. A. & D. L. Coplin, (1992) Exopolysaccharides in plant-bacterial interactions. *Annual review of microbiology* **46**: 307-346.

Leigh, J. A., E. R. Signer & G. C. Walker, (1985) Exopolysaccharide-deficient mutants of Rhizobium meliloti that form ineffective nodules. *Proceedings of the National Academy of Sciences of the United States of America* **82**: 6231-6235.

Levet-Paulo, M., J. C. Lazzaroni, C. Gilbert, D. Atlan, P. Doublet & A. Vianney, (2011) The atypical two-component sensor kinase Lpl0330 from Legionella pneumophila controls the bifunctional diguanylate cyclase-phosphodiesterase Lpl0329 to modulate bis-(3'-5')-cyclic dimeric GMP synthesis. *The Journal of biological chemistry* **286**: 31136-31144.

Li, Z., J. H. Chen, Y. Hao & S. K. Nair, (2012) Structures of the PelD cyclic diguanylate effector involved in pellicle formation in Pseudomonas aeruginosa PAO1. *The Journal of biological chemistry* **287**: 30191-30204.

Liang, Z. X., (2015) The expanding roles of c-di-GMP in the biosynthesis of exopolysaccharides and secondary metabolites. *Natural product reports* **32**: 663-683.

Lindstrom, K., M. Murwira, A. Willems & N. Altier, (2010) The biodiversity of beneficial microbe-host mutualism: the case of rhizobia. *Research in microbiology* **161**: 453-463.

Liu, T., C. F. Tian & W. X. Chen, (2015) Site-Specific Ser/Thr/Tyr Phosphoproteome of <italic>Sinorhizobium meliloti</italic> at Stationary Phase. *PloS one* **10**: e0139143.

Lynch, D., J. O'Brien, T. Welch, P. Clarke, P. O. Cuiv, J. H. Crosa & M. O'Connell, (2001) Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by Sinorhizobium meliloti. *Journal of bacteriology* **183**: 2576-2585.

Lloret, J., B. B. Wulff, J. M. Rubio, J. A. Downie, I. Bonilla & R. Rivilla, (1998) Exopolysaccharide II production is regulated by salt in the halotolerant strain Rhizobium meliloti EFB1. *Applied and environmental microbiology* **64**: 1024-1028.

Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng & S. H. Bryant, (2015) CDD: NCBI's conserved domain database. *Nucleic acids research* **43**: D222-226.

Margaret-Oliver, I., W. Lei, M. Parada, M. A. Rodriguez-Carvajal, J. C. Crespo-Rivas, A. Hidalgo, A. Gil-Serrano, J. Moreno, D. N. Rodriguez-Navarro, A. Buendia-Claveria, J. Ollero, J. E. Ruiz-Sainz & J. M. Vinardell, (2012) Sinorhizobium fredii HH103 does not strictly require KPS and/or EPS to nodulate Glycyrrhiza

uralensis, an indeterminate nodule-forming legume. *Archives of microbiology* **194**: 87-102.

Marketon, M. M., M. R. Gronquist, A. Eberhard & J. E. Gonzalez, (2002) Characterization of the Sinorhizobium meliloti sinR/sinI locus and the production of novel N-acyl homoserine lactones. *Journal of bacteriology* **184**: 5686-5695.

Matilla, M. A., M. L. Travieso, J. L. Ramos & M. I. Ramos-Gonzalez, (2011) Cyclic diguanylate turnover mediated by the sole GGDEF/EAL response regulator in Pseudomonas putida: its role in the rhizosphere and an analysis of its target processes. *Environmental microbiology* **13**: 1745-1766.

Matthysse, A. G., (2014) Attachment of Agrobacterium to plant surfaces. *Frontiers in plant science* **5**: 252.

Mattoo, S., M. H. Yuk, L. L. Huang & J. F. Miller, (2004) Regulation of type III secretion in Bordetella. *Molecular microbiology* **52**: 1201-1214.

Mijakovic, I. & B. Macek, (2012) Impact of phosphoproteomics on studies of bacterial physiology. *FEMS microbiology reviews* **36**: 877-892.

Miller, J. H., (1972) *Experiments in molecular genetics,* p. xvi, 466 p. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Miransari, M., (2011) Soil microbes and plant fertilization. *Applied microbiology and biotechnology* **92**: 875-885.

Mitchell, A., H. Y. Chang, L. Daugherty, M. Fraser, S. Hunter, R. Lopez, C. McAnulla, C. McMenamin, G. Nuka, S. Pesseat, A. Sangrador-Vegas, M. Scheremetjew, C. Rato, S. Y. Yong, A. Bateman, M. Punta, T. K. Attwood, C. J. Sigrist, N. Redaschi, C. Rivoire, I. Xenarios, D. Kahn, D. Guyot, P. Bork, I. Letunic, J. Gough, M. Oates, D. Haft, H. Huang, D. A. Natale, C. H. Wu, C. Orengo, I. Sillitoe, H. Mi, P. D. Thomas & R. D. Finn, (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic acids research* **43**: D213-221.

Morris, A. R., C. L. Darnell & K. L. Visick, (2011) Inactivation of a novel response regulator is necessary for biofilm formation and host colonization by Vibrio fischeri. *Molecular microbiology* **82**: 114-130.

Morris, A. R. & K. L. Visick, (2013a) Inhibition of SypG-induced biofilms and host colonization by the negative regulator SypE in Vibrio fischeri. *PloS one* **8**: e60076.

Morris, A. R. & K. L. Visick, (2013b) The response regulator SypE controls biofilm formation and colonization through phosphorylation of the syp-encoded regulator SypA in Vibrio fischeri. *Molecular microbiology* **87**: 509-525.

Najafi, S. M., D. A. Harris & M. D. Yudkin, (1996) The SpoIIAA protein of Bacillus subtilis has GTP-binding properties. *Journal of bacteriology* **178**: 6632-6634.

Najafi, S. M., A. C. Willis & M. D. Yudkin, (1995) Site of phosphorylation of SpoIIAA, the anti-anti-sigma factor for sporulation-specific sigma F of Bacillus subtilis. *Journal of bacteriology* **177**: 2912-2913.

Nakanishi, I., K. Kimura, T. Suzuki, M. Ishikawa, I. Banno, T. Sakane & T. Harada, (1976) Demonstration of curdlan-type polysaccharide and some other β-1,3-glucan in microorganisms with aniline blue. *The Journal of General and Applied Microbiology* **22**: 1-11.

Niehaus, K., D. Kapp & A. Pühler, (1993) Plant defence and delayed infection of alfalfa pseudonodules induced by an exopolysaccharide (EPS I)-deficient Rhizobium meliloti mutant. *Planta* **190**: 415-425.

Nogales, J., A. Dominguez-Ferreras, C. V. Amaya-Gomez, P. van Dillewijn, V. Cuellar, J. Sanjuan, J. Olivares & M. J. Soto, (2010) Transcriptome profiling of a Sinorhizobium meliloti fadD mutant reveals the role of rhizobactin 1021 biosynthesis and regulation genes in the control of swarming. *BMC genomics* **11**: 157.

Notredame, C., D. G. Higgins & J. Heringa, (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology* **302**: 205-217.

Oglesby, L. L., S. Jain & D. E. Ohman, (2008) Membrane topology and roles of Pseudomonas aeruginosa Alg8 and Alg44 in alginate polymerization. *Microbiology* **154**: 1605-1615.

Ohara, G. W., M. J. Dilworth, N. Boonkerd & P. Parkpian, (1988) Iron-deficiency specifically limits nodule development in peanut inoculated with *Bradyrhizobium sp*. *New Phytol.* **108**: 51-57.

Osterberg, S., A. Aberg, M. K. Herrera Seitz, M. Wolf-Watz & V. Shingler, (2013) Genetic dissection of a motility-associated c-di-GMP signalling protein of Pseudomonas putida. *Environmental microbiology reports* **5**: 556-565.

Pellock, B. J., H. P. Cheng & G. C. Walker, (2000) Alfalfa root nodule invasion efficiency is dependent on Sinorhizobium meliloti polysaccharides. *Journal of bacteriology* **182**: 4310-4318.

Pellock, B. J., M. Teplitski, R. P. Boinay, W. D. Bauer & G. C. Walker, (2002) A LuxR homolog controls production of symbiotically active extracellular polysaccharide II by Sinorhizobium meliloti. *Journal of bacteriology* **184**: 5067-5076.

Perez-Mendoza, D., I. M. Aragon, H. A. Prada-Ramirez, L. Romero-Jimenez, C. Ramos, M. T. Gallegos & J. Sanjuan, (2014) Responses to elevated c-di-GMP levels in mutualistic and pathogenic plant-interacting bacteria. *PloS one* **9**: e91645.

Perez-Mendoza, D., S. J. Coulthurst, S. Humphris, E. Campbell, M. Welch, I. K. Toth & G. P. Salmond, (2011) A multi-repeat adhesin of the phytopathogen, Pectobacterium atrosepticum, is secreted by a Type I pathway and is subject to complex regulation involving a non-canonical diguanylate cyclase. *Molecular microbiology* **82**: 719-733.

Perez-Mendoza, D., M. A. Rodriguez-Carvajal, L. Romero-Jimenez, A. Farias Gde, J. Lloret, M. T. Gallegos & J. Sanjuan, (2015) Novel mixed-linkage beta-glucan activated by c-di-GMP in Sinorhizobium meliloti. *Proceedings of the National Academy of Sciences of the United States of America* **112**: E757-765.

Pierson, E. A., D. W. Wood, J. A. Cannon, F. M. Blachere & L. S. Pierson, (1998) Interpopulation signaling via N-acyl-homoserine lactones among bacteria in the wheat rhizosphere. *Mol. Plant-Microbe Interact.* **11**: 1078-1084.

Pobigaylo, N., D. Wetter, S. Szymczak, U. Schiller, S. Kurtz, F. Meyer, T. W. Nattkemper & A. Becker, (2006) Construction of a large signature-tagged mini-Tn5 transposon library and its application to mutagenesis of Sinorhizobium meliloti. *Applied and environmental microbiology* **72**: 4329-4337.

Ramirez-Mata, A., I. J. Fernandez-Dominguez, K. J. Nunez-Reza, M. L. Xiqui-Vazquez & B. E. Baca, (2014) [Networks involving quorum sensing, cyclic-di-GMP and nitric oxide on biofilm production in bacteria]. *Revista Argentina de microbiologia* **46**: 242-255.

Ridge, R. W., (1993) A model of legume root hair-growth and *Rhizobium* infection. *Symbiosis* **14**: 359-373.

Rinaudi, L. V. & W. Giordano, (2010) An integrated view of biofilm formation in rhizobia. *FEMS microbiology letters* **304**: 1-11.

Rinaudi, L. V. & J. E. Gonzalez, (2009) The low-molecular-weight fraction of exopolysaccharide II from Sinorhizobium meliloti is a crucial determinant of biofilm formation. *Journal of bacteriology* **191**: 7216-7224.

Romling, U., (2002) Molecular biology of cellulose production in bacteria. *Research in microbiology* **153**: 205-212.

Romling, U., M. Y. Galperin & M. Gomelsky, (2013) Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiology and molecular biology reviews : MMBR* **77**: 1-52.

Ross, P., H. Weinhouse, Y. Aloni, D. Michaeli, P. Weinberger-Ohana, R. Mayer, S. Braun, E. de Vroom, G. A. van der Marel, J. H. van Boom & M. Benziman, (1987) Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. *Nature* **325**: 279-281.

Ryan, R. P. & J. M. Dow, (2010) Cell-cell signal dependent dynamic interactions between HD-GYP and GGDEF domain proteins mediate virulence in Xanthomonas campestris. *Virulence* **1**: 404-408.

Ryan, R. P., T. Tolker-Nielsen & J. M. Dow, (2012) When the PilZ don't work: effectors for cyclic di-GMP action in bacteria. *Trends in microbiology* **20**: 235-242.

Ryder, C., M. Byrd & D. J. Wozniak, (2007) Role of polysaccharides in Pseudomonas aeruginosa biofilm development. *Current opinion in microbiology* **10**: 644-648.

Sanchez-Contreras, M., W. D. Bauer, M. Gao, J. B. Robinson & J. Allan Downie, (2007) Quorum-sensing regulation in rhizobia and its role in symbiotic interactions with legumes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **362**: 1149-1163.

Schafer, A., A. Tauch, W. Jager, J. Kalinowski, G. Thierbach & A. Puhler, (1994) Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli

plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene* **145**: 69-73.

Schluter, J. P., J. Reinkensmeier, M. J. Barnett, C. Lang, E. Krol, R. Giegerich, S. R. Long & A. Becker, (2013) Global mapping of transcription start sites and promoter motifs in the symbiotic alpha-proteobacterium Sinorhizobium meliloti 1021. *BMC genomics* **14**: 156.

Schmid, J., V. Sieber & B. Rehm, (2015) Bacterial exopolysaccharides: biosynthesis pathways and engineering strategies. *Frontiers in microbiology* **6**: 496.

Schneiker, S., O. Perlova, O. Kaiser, K. Gerth, A. Alici, M. O. Altmeyer, D. Bartels, T. Bekel, S. Beyer, E. Bode, H. B. Bode, C. J. Bolten, J. V. Choudhuri, S. Doss, Y. A. Elnakady, B. Frank, L. Gaigalat, A. Goesmann, C. Groeger, F. Gross, L. Jelsbak, L. Jelsbak, J. Kalinowski, C. Kegler, T. Knauber, S. Konietzny, M. Kopp, L. Krause, D. Krug, B. Linke, T. Mahmud, R. Martinez-Arias, A. C. McHardy, M. Merai, F. Meyer, S. Mormann, J. Munoz-Dorado, J. Perez, S. Pradella, S. Rachid, G. Raddatz, F. Rosenau, C. Ruckert, F. Sasse, M. Scharfe, S. C. Schuster, G. Suen, A. Treuner-Lange, G. J. Velicer, F. J. Vorholter, K. J. Weissman, R. D. Welch, S. C. Wenzel, D. E. Whitworth, S. Wilhelm, C. Wittmann, H. Blocker, A. Puhler & R. Muller, (2007) Complete genome sequence of the myxobacterium Sorangium cellulosum. *Nature biotechnology* **25**: 1281-1289.

Sharma, A. K., A. C. Rigby & S. L. Alper, (2011) STAS domain structure and function. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* **28**: 407-422.

Simon, R., U. Priefer & A. Pühler, (1983) A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Nat Biotech* **1**: 784-791.

Skorupska, A., M. Janczarek, M. Marczak, A. Mazur & J. Krol, (2006) Rhizobial exopolysaccharides: genetic control and symbiotic functions. *Microbial cell factories* **5**: 7.

Sondermann, H., N. J. Shikuma & F. H. Yildiz, (2012) You've come a long way: c-di-GMP signaling. *Current opinion in microbiology* **15**: 140-146.

Sorokin, D. Y., T. P. Tourova, A. M. Lysenko, L. L. Mityushina & J. G. Kuenen, (2002) Thioalkalivibrio thiocyanoxidans sp. nov. and Thioalkalivibrio paradoxus sp. nov., novel alkaliphilic, obligately autotrophic, sulfur-oxidizing bacteria capable of growth on thiocyanate, from soda lakes. *International journal of systematic and evolutionary microbiology* **52**: 657-664.

Sorroche, F. G., L. V. Rinaudi, A. Zorreguieta & W. Giordano, (2010) EPS II-dependent autoaggregation of Sinorhizobium meliloti planktonic cells. *Current microbiology* **61**: 465-470.

Srivastava, D., R. C. Harris & C. M. Waters, (2011) Integration of cyclic di-GMP and quorum sensing in the control of vpsT and aphA in Vibrio cholerae. *Journal of bacteriology* **193**: 6331-6341.

Srivastava, D. & C. M. Waters, (2012) A tangled web: regulatory connections between quorum sensing and cyclic Di-GMP. *Journal of bacteriology* **194**: 4485-4493.

Stanley, N. R. & B. A. Lazazzera, (2004) Environmental signals and regulatory pathways that influence biofilm formation. *Molecular microbiology* **52**: 917-924.

Stasinopoulos, S. J., P. R. Fisher, B. A. Stone & V. A. Stanisich, (1999) Detection of two loci involved in (1-->3)-beta-glucan (curdlan) biosynthesis by Agrobacterium sp. ATCC31749, and comparative sequence analysis of the putative curdlan synthase gene. *Glycobiology* **9**: 31-41.

Steiner, S., C. Lori, A. Boehm & U. Jenal, (2013) Allosteric activation of exopolysaccharide synthesis through cyclic di-GMP-stimulated protein-protein interaction. *The EMBO journal* **32**: 354-368.

Sutherland, I. W., (2001) The biofilm matrix - an immobilized but dynamic microbial environment. *Trends in microbiology* **9**: 222-227.

Tal, R., H. C. Wong, R. Calhoon, D. Gelfand, A. L. Fear, G. Volman, R. Mayer, P. Ross, D. Amikam, H. Weinhouse, A. Cohen, S. Sapir, P. Ohana & M. Benziman, (1998) Three cdg operons control cellular turnover of cyclic di-GMP in Acetobacter xylinum: genetic organization and occurrence of conserved domains in isoenzymes. *Journal of bacteriology* **180**: 4416-4425.

Tarutina, M., D. A. Ryjenkov & M. Gomelsky, (2006) An unorthodox bacteriophytochrome from Rhodobacter sphaeroides involved in turnover of the second messenger c-di-GMP. *The Journal of biological chemistry* **281**: 34751-34758.

Telford, G., D. Wheeler, P. Williams, P. T. Tomkins, P. Appleby, H. Sewell, G. S. Stewart, B. W. Bycroft & D. I. Pritchard, (1998) The Pseudomonas aeruginosa quorum-sensing signal molecule N-(3-oxododecanoyl)-L-homoserine lactone has immunomodulatory activity. *Infection and immunity* **66**: 36-42.

Thompson, C. C., C. Griffiths, S. S. Nicod, N. M. Lowden, S. Wigneshweraraj, D. J. Fisher & M. O. McClure, (2015) The Rsb Phosphoregulatory Network Controls Availability of the Primary Sigma Factor in Chlamydia trachomatis and Influences the Kinetics of Growth and Development. *PLoS pathogens* **11**: e1005125.

Torres, M. J., M. I. Rubia, E. J. Bedmar & M. J. Delgado, (2011) Denitrification in Sinorhizobium meliloti. *Biochemical Society transactions* **39**: 1886-1889.

Torres, M. J., M. I. Rubia, T. C. de la Pena, J. J. Pueyo, E. J. Bedmar & M. J. Delgado, (2014) Genetic basis for denitrification in Ensifer meliloti. *BMC microbiology* **14**: 142.

Trampari, E., C. E. Stevenson, R. H. Little, T. Wilhelm, D. M. Lawson & J. G. Malone, (2015) Bacterial Rotary Export ATPases Are Allosterically Regulated by the Nucleotide Second Messenger Cyclic-di-GMP. *The Journal of biological chemistry* **290**: 24470-24483.

van Workum, W. A. T., S. van Slageren, A. A. N. van Brussel & J. W. Kijne, (1998) Role of exopolysaccharides of Rhizobium leguminosarum bv. viciae as host plant-specific molecules required for infection thread formation during nodulation of Vicia sativa. *Mol. Plant-Microbe Interact.* **11**: 1233-1241.

Vaningelgem, F., M. Zamfir, F. Mozzi, T. Adriany, M. Vancanneyt, J. Swings & L. De Vuyst, (2004) Biodiversity of exopolysaccharides produced by Streptococcus thermophilus strains is reflected in their production and their molecular and functional characteristics. *Applied and environmental microbiology* **70**: 900-912.

Viguier, C., O. C. P, P. Clarke & M. O'Connell, (2005) RirA is the iron response regulator of the rhizobactin 1021 biosynthesis and transport genes in Sinorhizobium meliloti 2011. *FEMS microbiology letters* **246**: 235-242.

Vinuesa, P., (2015) Rhizobial Taxonomy Up-to-Date [Online]. Mexico: Universidad Nacional Autonoma de Mexico; [accessed November, 2015].

Visick, K. L., (2009) An intricate network of regulators controls biofilm formation and colonization by Vibrio fischeri. *Molecular microbiology* **74**: 782-789.

Wang, H. & A. Norén, (2006) Metabolic regulation of nitrogen fixation in Rhodospirillum rubrum. *Biochemical Society transactions* **34**: 160-161.

Wang, Y., W. Lin, J. Li & Y. Pan, (2013) Changes of cell growth and magnetosome biomineralization in Magnetospirillum magneticum AMB-1 after ultraviolet-B irradiation. *Frontiers in microbiology* **4**: 397.

Waters, C. M., W. Lu, J. D. Rabinowitz & B. L. Bassler, (2008) Quorum sensing controls biofilm formation in Vibrio cholerae through modulation of cyclic di-GMP levels and repression of vpsT. *Journal of bacteriology* **190**: 2527-2536.

Wexler, M., K. H. Yeoman, J. B. Stevens, N. G. de Luca, G. Sawers & A. W. Johnston, (2001) The Rhizobium leguminosarum tonB gene is required for the uptake of siderophore and haem as sources of iron. *Molecular microbiology* **41**: 801-816.

Whitfield, G. B., L. S. Marmont & P. L. Howell, (2015) Enzymatic modifications of exopolysaccharides enhance bacterial persistence. *Frontiers in microbiology* **6**: 471.

Wigren, E., Z. X. Liang & U. Romling, (2014) Finally! The structural secrets of a HD-GYP phosphodiesterase revealed. *Molecular microbiology* **91**: 1-5.

Williams, A., A. Wilkinson, M. Krehenbrink, D. M. Russo, A. Zorreguieta & J. A. Downie, (2008) Glucomannan-mediated attachment of Rhizobium leguminosarum to pea root hairs is required for competitive nodule infection. *Journal of bacteriology* **190**: 4706-4715.

Williams, P., (2007) Quorum sensing, communication and cross-kingdom signalling in the bacterial world. *Microbiology* **153**: 3923-3938.

Williams, P., K. Winzer, W. C. Chan & M. Camara, (2007) Look who's talking: communication and quorum sensing in the bacterial world. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **362**: 1119-1134.

Wood, P. J., (1980) SPECIFICITY IN THE INTERACTION OF DIRECT DYES WITH POLYSACCHARIDES. *Carbohydrate Research* **85**: 271-287.

Yang, C. Y., K. H. Chin, M. L. Chuah, Z. X. Liang, A. H. Wang & S. H. Chou, (2011) The structure and inhibition of a GGDEF diguanylate cyclase complexed with (c-di-

GMP)(2) at the active site. *Acta crystallographica. Section D, Biological crystallography* **67**: 997-1008.

Yang, X., C. M. Kang, M. S. Brody & C. W. Price, (1996) Opposing pairs of serine protein kinases and phosphatases transmit signals of environmental stress to activate a bacterial transcription factor. *Genes & development* **10**: 2265-2275.

Yudkin, M. D. & J. Clarkson, (2005) Differential gene expression in genetically identical sister cells: the initiation of sporulation in Bacillus subtilis. *Molecular microbiology* **56**: 578-589.

Zhang, X. S. & H. P. Cheng, (2006) Identification of Sinorhizobium meliloti early symbiotic genes by use of a positive functional screen. *Applied and environmental microbiology* **72**: 2738-2748.

Zhou, Q., P. Ames & J. S. Parkinson, (2009) Mutational analyses of HAMP helices suggest a dynamic bundle model of input-output signalling in chemoreceptors. *Molecular microbiology* **73**: 801-814.